*Second Edition*

# MICROELECTRONICS

*Jacob Millman*
*Arvin Grabel*






**MICROELECTRONICS**



**Tata McGraw-Hill Edition 1999**

22nd reprint 2009
**RAZDCDLXDRRRD**

Reprinted in India by arrangement with The McGraw-Hill Companies, Inc., New York

**For Sale in India Only**

**ISBN-13: 978-0-07-463736-4**
**ISBN-10: 0-07-463736-3**

Published by Tata McGraw-Hill Publishing Company Limited, 7 West Patel Nagar, New Delhi 110 008, and printed at Gopaljee Enterprises, Delhi 110 053

*The McGraw·Hill Companies*

# CONTENTS

provides the necessary background for either section. Our reasons for introducing digital material first are twofold:

1. In many universities, computer engineering and science majors are required to take only one electronics course. Clearly, for these students such a course should concentrate on digital electronics. Judicious selection of topics in Parts 1 and 2 can create a satisfactory one-term course, and this opportunity gives the instructor the freedom and the incentive to consider the subject matter most appropriate to his or her objectives.
2. Only elementary circuit theory of the level described in Appendix C is required. Consequently, the student need not have any electrical engineering prerequisites to handle these topics.

The first chapter in Part 2 treats the basic logic-gate building blocks used in digital systems. We concentrate on the operation and performance of the four major IC technologies. The two field-effect transistor (FET) families are NMOS and CMOS, whereas transistor-transistor logic (TTL) and emitter-coupled logic (ECL) are the standard bipolar junction transistor (BJT) logic families investigated. The standard small-scale integration (SSI), medium-scale integration (MSI), and large-scale integration (LSI) circuits and systems derived from these logic gates are developed in the next two chapters (combinatorial and sequential circuits and systems). The last chapter deals with very large scale integration (VLSI) systems in which static and dynamic random-access memory (RAM) cells are discussed. Technologies used only in VLSI systems, such as integrated-injection logic ($I^2L$), CMOS domino logic, and charge-coupled devices (CCD), are also introduced.

The development in Part 3 (Chapters 10 to 14) on *amplifier circuits and systems* parallels that in Part 2. The initial pair of chapters focuses on the properties of basic BJT and FET amplifier stages. Methods by which IC transistors are biased are presented; the use of small-signal models to evaluate the performance of single-stage and cascaded amplifiers is discussed. The operational amplifier (Op-Amp) as a basic building block is also described. These amplifiers are used to form feedback-amplifier systems. The fundamental feedback concepts and techniques developed are used in the analysis and design of the four basic feedback-amplifier topologies. The internal architecture and performance of modern Op-Amp chips are examined in Chapter 14 and serve to bring together many of the concepts previously encountered in Part 3.

Part 4 (Chapters 15 and 16) examines the circuit and systems aspect of *data acquisition and signal processing*. Many of these circuits are used in both digital and analog systems and employ both logic gates and Op-Amps. Circuits for the generation of sinusoidal, ramp, and pulse waveforms and the conversion of analog signals to digital signals (or vice versa) are discussed. Integrators, active filters including switched-capacitor filters, multipliers, and logarithmic amplifiers are among the signal-conditioning circuits described.

The last part (Chapter 17) exposes the student to electronic *power circuits*

*and systems*. The conversion of ac to dc is treated and leads to a discussion of monolithic voltage regulators. High-voltage and high-power amplifier devices and circuits are also examined.

This text contains enough material for 2 or 3 one-semester electronics devices-circuits-systems courses. With the ever-increasing component density on an IC chip, the difference between an electronic device, circuit, and system has become quite blurred, and in this book no attempt is made to distinguish between them. An entire monolithic package, such as an Op-Amp, is often referred to as a *device*. Of course, a single transistor is clearly a *device* and a large-scale microelectronic chip merits the designation *system* or at least *subsystem*.

A brief historical survey of electronics is in the Prologue (following this preface). It is hoped that both the instructor and student will read this fascinating history before beginning the study of the text.

Most electronic engineers design a new product, subsystem, or system by interconnecting standard IC chips so that the overall assembly achieves the desired external objectives. Clearly these engineers must know what IC chips are commercially available, what function they perform, and what their limitations are. Chip designers must be aware of what functions need to be performed and what limitations most affect the performance of the systems in which the chips are used.

From this perspective, the goal of this book is to take the reader step by step from a qualitative knowledge of semiconductor properties to an understanding of the operation of solid-state devices and finally to an appreciation for how these are combined to form ICs with distinct and useful input-output characteristics. A very broad variety of IC chips are studied in this book. We describe not only what is fabricated, but also attempt to convey a deep understanding of the digital and/or analog functions performed by the chip. After each circuit or system is studied, reference is made to a specific commercially available chip which realizes the desired function. Practical limitations of real rather than ideal devices and circuits are explained. To appreciate nonideal behavior, manufacturer's specifications of representative devices and integrated circuits are given in Appendix B. The depth of discussion, the broad choice of topics, and the practical emphasis combine to prepare the student to do useful engineering immediately upon graduation.

The attention given to pedagogy is reflected in the explanation of device-circuit-system behavior and in the context in which the specific topics are discussed. We have been diligent in our efforts to ensure that new concepts are introduced by the use of familiar analytic techniques and that the development of new methods of analysis relies only on concepts previously encountered. Also, considerable care was exercised in the selection of the many illustrative examples and numerical calculations incorporated in the body of the text.

Many of the methods of analysis discussed lead to the "pencil-and-paper" calculations which an engineer often performs. Such computations are invaluable as they help develop insight into the behavior of the circuit or system

being designed. When used in conjunction with computer simulations, they provide the engineer with powerful design tools. Students should be encouraged to use circuit simulations such as SPICE and MICROCAP II, both of which are available for use with personal computers.

The review questions at the end of each chapter are a significant adjunct to the approximately 800 problems given in Appendix D. Many of these problems are new, and the majority of the problems used in the earlier edition have been modified. Used together, the questions and problems test the students' grasp of fundamental concepts and provide experience in the design and analysis of electronic circuits. Realistic parameter values are used in virtually all numerical problems.

The review questions test the students' qualitative knowledge of the text material. These can also be used very effectively as part of a quiz or an exam.

A solutions manual is available to an instructor who has adopted the text. As an added pedagogical aid, transparencies of over 100 involved figures in the book are also available to the instructor.

We have had the benefit of valuable advice and suggestions from the many professors and engineers working in industry who used the first edition as either instructor, student, or practicing engineer. All these individuals have influenced this edition and to them, we express our thanks and appreciation. We are especially grateful to Professor Arthur Dickerson whose comments and insight were invaluable in the preparation of this book.

We are indebted to David Damstra, who from manuscript through production contributed much to this book; to the comments of and reviews provided by Sanjeev Rao; and to Mary Rosenberg, whose proofing of final pages was a great help.

**Jacob Millman**
**Arvin Grabel**

P.S. I am among the two generations of electrical engineers who have studied electronics from Jacob Millman's books. I had the pleasure of once again being Professor Millman's "student" when we worked closely together in the planning and organization of this book and in the detailed preparation of the first six chapters. I have attempted to convey the guiding spirit of this truly remarkable teacher and writer in the remainder of this text. The last eleven chapters were my sole responsibility and therefore a reflection of the quality of the student and not the mentor.

I am indebted to Jacob Millman for the opportunity of collaborating with him. His influence and style have contributed immeasurably to me as teacher and author.

**Arvin Grabel**

**THE VACUUM-TUBE ERA** The vacuum-tube era spans the first half of the twentieth century; modern electronics took shape technologically in this period.

The origin of the term "electronics" can be attributed to H. A. Lorentz, who in 1895 postulated the existence of discrete charges he called *electrons* (reintroducing the word for "amber" used by the ancient Greeks). Two years later J. J. Thompson experimentally verified the existence of electrons. In that same year Braun built the first electron tube, a primitive cathode-ray tube (CRT).

**The Discovery of Vacuum Tubes** In 1904 Fleming invented a two-element device, the diode, which he called the *valve*. It consisted of a heated wire, the filament, which emitted electrons (the Edison effect) and was separated by a short distance from a metallic plate. The entire structure was encapsulated in a vacuum. A positive plate-to-cathode (filament) voltage produced a current, whereas a negative applied voltage reduced the current to zero. This unilateral property of the valve made it useful as a detector of wireless (radio) signals.

Two years later, Pickard used a silicon crystal with a "cat's whisker" (a pointed wire pressed into the silicon) as a detector. This was the first semiconductor diode; however, it was unreliable and was soon abandoned. Thus semiconductor electronics appeared to have died a premature death in 1906.

The invention of the *audion* (triode) by DeForest in 1906 was the seminal achievement in the earliest days of electronics. Indeed, one can strongly argue that electronics as we know it today would not exist without the invention of the triode. DeForest's audion consisted of a third electrode (the grid) inserted between the plate and the cathode of the Fleming valve. The grid voltage controlled the charge flow between plate and cathode. A small change in grid voltage resulted in a larger plate-voltage change, making the audion the first amplifier.

The triode was the first device to exhibit the circuit property we now refer to as a *controlled* or *dependent source*. Because it retained the unilateral property of the valve, the triode also provided the properties of a controlled switch. Today, virtually all electronic circuits exploit device characteristics which display either controlled-source or controlled-switch behavior.

**Initial Circuit Applications** By 1911, technological improvements—a better vacuum and an oxide-coated cathode—made the audion a reliable device, thus ushering in the age of practical electronics.[1] The first applications of vacuum tubes were to telephone and radio communication, and simultaneously the Institute of Radio Engineers (IRE) was founded in 1912 in the United States. It is a tribute to the imagination and foresight of the early engineers who immediately realized the significance of radio and formed their own professional society. The American Institute of Electrical Engineers (AIEE), which focused on the conventional

---

[1] By coincidence, Professor Millman was born the same year.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

**Controls**

The origin of the electronic control industries was in "industrial electronics," which may be defined as "the use of electronic devices in the control of machines (other than in communication and computation)." Thyratrons, gaseous diodes, mercury arc rectifiers, and high-voltage high-power tubes were the devices used. These devices were used in circuits which provided high-voltage and high-power alternating-current-to-direct-current (ac-to-dc) conversion (rectifiers), dc-to-ac conversion (inverters), and high-voltage transmitting circuits. Applications included motor-speed control, voltage regulation, induction and dielectric heating, and a variety of industrial process controls. Also, the first use of computers (analog) in control systems appeared at this time.

**Analyses and Theory** In addition to industrial growth, significant analytical and theoretical progress was made. The following is a brief indication of the range of achievements.

Circuit analysis and synthesis techniques were developed, notably by groups at Bell Laboratories and MIT. Bode and Nyquist developed feedback amplifier theory and transformed Black's circuit concept into one enjoying widespread use and import.

Shannon in the United States and Kotelnikov in the Soviet Union independently developed information theory which was to have great impact on data transmission. One particular application was to the pulse-code-modulation (PCM) technique proposed by Reeves.

The use of boolean algebra in the analysis and design of switching circuits was another of Shannon's contributions (1937). In Great Britain, the concept of a universal computing machine was proposed by Turing, and Wilkes developed microprogramming.

Sampled-data systems, introduced by Ragazzini and Zadeh, were applied in control applications, paving the way for control systems based on digital-computer processing.

Studies of materials, particularly the application of quantum mechanics to solids, led to new devices and later helped to lead to the invention of the transistor. Transducers, devices by which light, sound, pressure, temperature, and other variables are converted to and from electrical signals, were introduced to make use of the advantages afforded by electronics.

New forms of instrumentation (oscilloscopes, vacuum-tube voltmeters, etc.) evolved to both use electronics in measurement and for testing electronic equipment.

The 1950s was a decade of transition. It marked the end of the development of sophisticated vacuum-tube systems and the beginning of the transistor age. Today, the entire field is dominated by semiconductor devices except for high-voltage high-power applications. Indeed, vacuum tubes are omitted from virtually all electrical engineering curricula.

not under laboratory conditions. The following approximate dates give some indication of the increasing component count per chip:

1951—*discrete transistors*
1960—small-scale integration (SSI), fewer than 100 components
1966—medium-scale integration (MSI), 100 to 1000 components
1969—large-scale integration (LSI), 1000 to 10,000 components
1975—very-large-scale integration (VLSI),[1] more than 10,000 components

One can divide the electronic industries into chip manufacturers and chip users. The IC manufacturers are the major segment of the component industries, whereas chip users are most often the companies producing communication, control, and computer equipment. Since the invention of the IC, many innovations have contributed to the growth of microelectronics. Several of these are described in the remainder of this section.

**The Field-Effect Transistor** Much of the work leading to the invention of the bipolar transistor involved studies of the effect an applied electric field had on the conductivity of semiconductors. Shockley proposed the junction field-effect transistor (JFET) in 1951, but early attempts at fabrication failed because a stable surface could not be obtained. This difficulty was overcome with the introduction of the planar process and silicon dioxide ($SiO_2$) passivation. In 1958, the first JFET was produced by Teszner in France.

The techniques used to make reliable JFETs led to an even more important device, the metal-oxide-semiconductor field-effect transistor (MOSFET). The structure consists of a metallic electrode (the gate) placed on the $SiO_2$ between two electrodes in the semiconductor (source and drain). The current in the "channel" between source and drain can be controlled by applying an appropriate voltage between the gate and the semiconductor. Atalla and Kahng (1960) at Bell Laboratories reported the first such device. Two years later, Hofstein and Heiman at RCA were awarded a patent for their development of MOSFETs suitable for IC fabrication. Subsequent improvements in processing and device design and the growth of the computer industry have made MOS devices the most widely used transistors.

**Digital Integrated Circuits** The growth of the computer industry spurred new IC development; in turn, new IC concepts resulted in new computer architectures. Two of the major advances were in new circuit configurations and semiconductor memories.

Speed, power consumption, and component density are important considerations in digital ICs. An early bipolar logic family was transistor-coupled transistor logic, invented by Buie (1961) of Pacific Semiconductor,[2] from which

---

[1]By 1984, most VLSI chips had 100,000 or more components.

[2]Pacific Semiconductor is now part of TRW.

Currently, in the 1980s, the fourth generation of machines is being developed and introduced. These computers employ VLSI chips for both processing and memory. Today, electronic computers are available in a variety of sizes ranging from the simplest of microprocessors to supercomputers capable of executing tens of millions of instructions per second.

To achieve higher speeds, increased computational capability, and more flexible processing, many innovations were made. In addition to faster high-density chips, these include parallel processing, pipelining, and new concepts in compilers and assemblers. In addition, time-sharing and distributed computation have had an important effect on computer usage.

The impact of microelectronics was dramatically expressed by Noyce in 1977: "Today's microcomputer at a cost of perhaps $300 has more computing capacity than the first large electronic computer, ENIAC. It is 20 times faster, has a larger memory, is thousands of times more reliable, consumes the power of a light bulb rather than that of a locomotive, occupies 1/30,000 the volume and costs 1/10,000 as much. It is available by mail order or from your local hobby shop."

## THE FUTURE

Throughout most of your (the readers') lives, it has been possible to have live television communication from anywhere on Earth or, in fact, from millions of miles in space. What is amazing is not that these are nearly everyday events, but that they can be achieved at all. It is an awesome accomplishment when we consider that someone at the Johnson Space Center can throw a switch and instruct a space vehicle 1 *billion* kilometers (km) away to turn on its TV camera, focus it, and send pictures back to Earth. (Even at the speed of light it takes nearly 2 hours to transmit the instruction and receive the signals.) None of this is possible without the advances in electronics, culminated by the IC, that were described in the two previous sections. However, even as this is history, the achievement is indicative of the future direction of electronics.

The ability to transmit television pictures from a spacecraft requires that communications, computation, and control equipment act in unison as a single entity. It is apparent that the areas of electronics are merging and the "intelligent" electronic systems that result are at the core of the information age.[1]

The marriage of extensive communications and inexpensive computers has already begun to penetrate nearly every aspect of society. In addition to traditional industrial applications, the ability and relative ease by which information can be stored, retrieved, manipulated, and transmitted has affected us in our homes, our places of work, and our means for getting from one to the other. Office automation (word processors, electronic mail, etc.) is transforming how and where we work. Energy management, appliance control, security

[1]The era of the 1980s extending into the twenty-first century has been called the "information age," partly because more than 50 percent of the U.S. workforce can be classified as "information workers."

systems, cable television, and personal computers are some microelectronic applications in the home. The computer-controlled Bay Area Rapid Transit (BART) system in metropolitan San Francisco and electronic ignition, emission control, and safety systems in automobiles are examples of the impact of electronics on transportation. So widespread will the impact of microelectronics be that Noyce has used the analogy that by the end of this century electronics will be like the electric motor of today—largely unnoticed.

We believe that the electronics industries will continue to be the four C's—components, communications, computation, and control. It will be exceedingly difficult to recognize them as separate entities because they will be merged to an even greater degree. Similarly, the distinction between device, circuit, and system will become increasingly blurred. For the next decade, silicon-based technology will dominate electronics. However, the results of research on new materials, particularly gallium arsenide (GaAs), will, in all probability, begin to play a significant role.[1]

The future impact of microelectronics is readily apparent from the following statistics and projections for the U.S. electronics market (in *billions* of dollars):

| | *1985* | *1990* |
|---|---|---|
| Electronic sales | 215 | 400 |
| Integrated circuit sales | 11 | 35 |

These projections indicate that the creativity and ingenuity of the engineers and scientists of the past will be the springboard for the engineering talent of the future.

## GENERAL REFERENCES


1 Fiftieth Anniversary Issue: *Electronics*, vol. 53, no. 9, April 1980.

2 Fiftieth Anniversary Issue: *Proceedings of the IRE*, vol. 50, no. 5, May 1962.

3 Special Issue: Historical Notes on Important Tubes and Semiconductor Devices, *IEEE Trans. Electron Devices*, vol. ED-23, no. 7, July 1976.

4 Weiner, C.: How the Transistor Emerged, *IEEE Spectrum*, pp. 24–33, January 1973.

5 Forester, T.: "The Microelectronics Revolution", MIT Press, Cambridge, Mass., 1981.

6 Special Issue: Microelectronics, *Scientific American*, September 1977.

7 Wolff, M. F.: The Genesis of the Integrated Circuit, *IEEE Spectrum*, pp. 45–53, August 1976.

8 Mayo, J. S.: Technical Requirements of the Information Age, *Bell Labs Record*, vol. 60, no. 55, 1982.

9 Kidder, T.: *The Soul of a New Machine*, Little, Brown and Company, Boston, 1981.


[1]There is even speculation that organic materials, such as DNA, may find use in electronics by the end of this century.

# *Part One*
# SEMICONDUCTOR DEVICES

Semiconductor devices are the central components used to process the electrical signals that arise in communication, computation, and control systems. It is the electrical behavior of these devices that provides the controlled sources and controlled switches needed in signal-processing circuits. In the five chapters of this section, the physical operation and characteristics of the major semiconductor devices are developed. Elementary circuit applications are introduced to illustrate how the device characteristics are exploited in switches and amplifiers. Chapter 1 deals with the concepts which govern the electrical properties of semiconductors. Chapters 2 to 4 treat junction diodes and bipolar and field-effect transistors. An overview of integrated-circuit (IC) fabrication is presented in Chap. 5.

# Chapter 1
# SEMICONDUCTORS

The controlled flow of charged particles is fundamental to the operation of all electronic devices. Consequently, the materials used in these devices must be capable of providing a source of mobile charges and the processes which govern the flow of charges must be amenable to control. The physical properties of semiconductors as they relate to electronic devices are described in this chapter. In particular, the electrical characteristics of materials which allow us to distinguish semiconductors from insulators and conductors and the use of doping a semiconductor with impurities to control its electrical behavior are discussed.

Two principal charge-transport processes are also investigated: (1) *drift*, which is the motion of charges produced by an electric field, and (2) *diffusion*, which is motion resulting from a nonuniform charge distribution.

## 1-1 FORCES, FIELDS, AND ENERGY

In this section we introduce the basic quantities used to describe the effects of charged particles. For most students, it is a brief review of material previously treated in physics courses.

**Charged Particles** The *electron* is the principal negatively charged particle whose *charge*, or quantity of electricity, has been determined as $1.60 \times 10^{-19}$ coulombs (C).[1] The number of electrons per coulomb is the reciprocal of the electronic charge or approximately $6 \times 10^{18}$. Since a current of 1 ampere (A) is 1 coulomb per second (C/s), a current of 1 picoampere (pA, or $10^{-12}$ A) represents the motion

[1]International System of Units (SI) units are used almost exclusively in this text. Abbreviations for units are used with numeric values and after symbolic equations; otherwise they are spelled out. Units, technical terms, and their abbreviations are listed in App. A-1. The values of many important physical constants are given in App. A-2, and conversion factors and prefixes are listed in App. A-3.

of approximately 6 million electrons. Yet a current of 1 pA is so small that considerable difficulty is experienced in attempting its measurement.

In dealing with atoms, it is often convenient to consider the positive nucleus and inner electronic bands as an equivalent positive charge (the ionic core) whose magnitude is an integral multiple of the charge on an electron. The number of *valence electrons*, that is, those in the outermost electronic band, provides a negative charge to make the atom neutral. Under certain conditions, one or more valence electrons may be removed from the atom, leaving a positive ion. Similarly, one or more electrons can be added to the valence band, creating a negative ion. For example, the sodium and chlorine ions in common salt are singly ionized particles, with each having a magnitude of charge equal to that of the electron. The sodium ion is positive and the chlorine ion negative, resulting from the removal and addition, respectively, of one valence electron. Doubly ionized particles have ionic charges equal to twice that of the electron.

In a silicon crystal, each silicon ion shares a pair of electrons with each of four ionic neighbors. This configuration is called a *covalent bond*. Circumstances may cause an electron to be missing from this structure, leaving a "hole" in the bond.[1] These vacancies may move from ion to ion in the crystal and produce an effect equivalent to that of the motion of positive charges. The magnitude of charge associated with the hole is that of an electron.

**Field Intensity**

In the vicinity of a charged particle an electric field is said to exist; that is, a charged particle exerts a force on other charged particles as given by Coulomb's law. For the one-dimensional case[2] in which charge $q_1$ is at $x_0$, the force exerted on charge $q_2$ at an arbitrary distance $x$ in newtons (N) is

$$F_x = \frac{q_1 q_2}{4\pi\epsilon(x - x_0)^2} \quad \text{N} \tag{1-1}$$

where $\epsilon$ is the permittivity of the medium in which the charges reside. From Newton's third law, an equal but opposite force acts on $q_1$.

The motion of $q_2$ is obtained by applying Newton's second law and is

$$F_x = \frac{q_1 q_2}{4\pi\epsilon(x - x_0)^2} = \frac{d}{dt}(m_2 v_x) \quad \text{N} \tag{1-2}$$

in which $m_2$ is the mass of $q_2$ and $v_x$ is its velocity in the $x$ direction. For a nonrelativistic system ($m_2$ is constant), Eq. (1-2) reduces to

$$F_x = m_2 \frac{dv_x}{dt} = m_2 a_x \quad \text{N} \tag{1-3}$$

where $a_x = dv_x/dt$ is the acceleration.

---

[1]This brief introduction to the hole as a charge carrier is discussed further in Sec. 1-3.

[2]The electric field and the forces it exerts are three-dimensional in general. However, in many electronic structures a uniform cross section permits one-dimensional representation.

A convenient method for describing the effect of charged particles is the use of the *electric field intensity* $\mathscr{E}$, defined as the force exerted on a unit positive charge. Thus the force on a charge $q$ in an electric field is, in one dimension,

$$F_x = q\mathscr{E}_x \qquad \text{N} \tag{1-4}$$

**Potential**

By definition, the potential $V$ [in volts (V)] of point $B$ with respect to point $A$ is the work done against the field in moving a unit positive charge from $A$ to $B$. In one dimension, with $A$ at $x_0$ and $B$ at an arbitrary distance $x$, it follows that[1]

$$V \equiv -\int_{x_0}^{x} \mathscr{E}_x \, dx \qquad \text{V} \tag{1-5}$$

where $\mathscr{E}_x$ represents the $x$ component of the field. Differentiation of Eq. (1-5) gives

$$\mathscr{E} = -\frac{dV}{dx} \qquad \text{V/m} \tag{1-6}$$

The minus sign indicates that the electric field is directed from the region of higher potential to the region of lower potential.

By definition, the potential energy $U$ equals the potential multiplied by the charge $q$ under consideration, or [in joules (J)]

$$U \equiv qV \qquad \text{J} \tag{1-7}$$

If an electron is being considered, $q$ is replaced by $-q$ (where $q$ is the magnitude of the electronic charge).

Because the energy associated with a single electron is so small, it is convenient to introduce the *electron volt* (eV) unit of energy (work), defined as

$$1 \text{ eV} = 1.60 \times 10^{-19} \text{ J}$$

Of course, any type of energy, whether it is electrical, mechanical, thermal, or similar, may be expressed in electron volts.

Equation (1-7) indicates that if an electron falls through a potential of 1 V, its kinetic energy increases and its potential energy decreases by $1.60 \times 10^{-19}$ J or 1 eV. Although each electron possesses a small amount of energy, an enormous number of electrons is required for even a small current. Consequently, electron devices can handle reasonable amounts of power.

The law of conservation of energy states that the total energy $W$, which equals the sum of the potential energy $U$ and the kinetic energy $\frac{1}{2}mv^2$, remains constant. Thus at any point

$$W = U + \tfrac{1}{2}mv^2 = \text{const} \tag{1-8}$$

---

[1] The symbol ≡ is used to designate "equal to by definition."

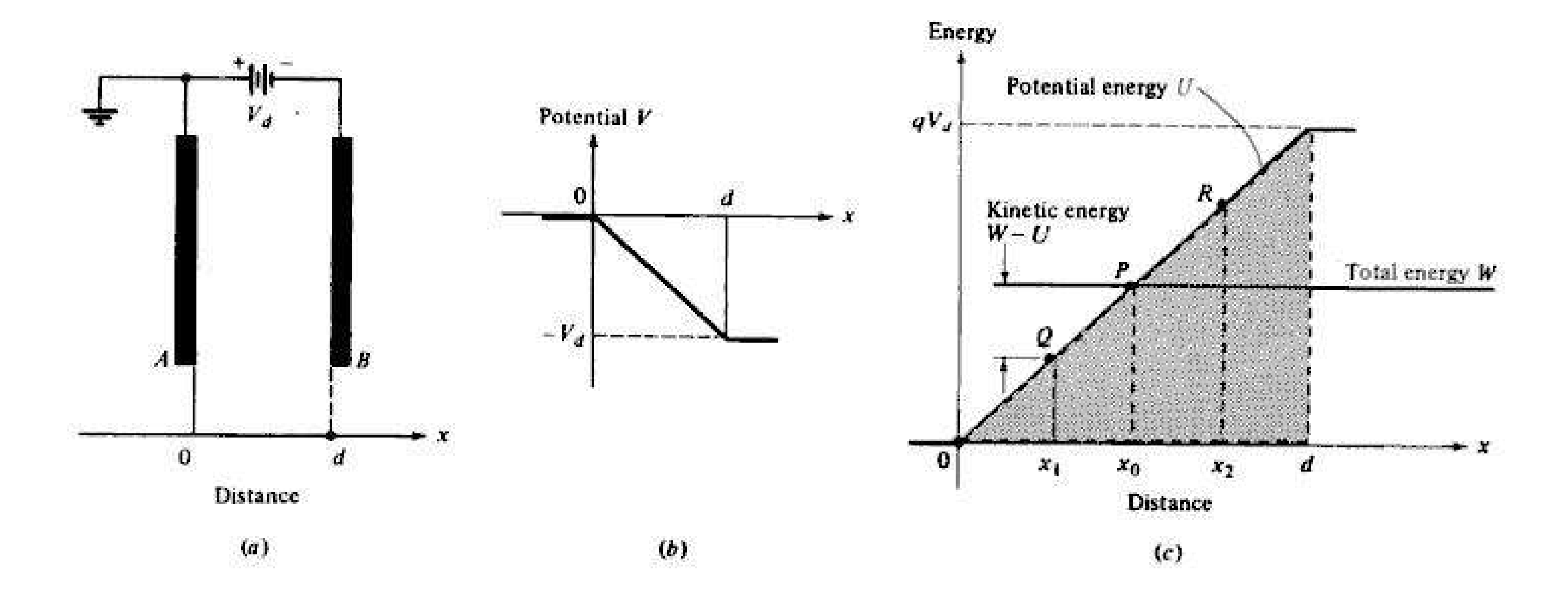



**FIGURE 1-1**
(*a*) Parallel-plate system showing an electron leaving *A* with an initial velocity $v_0$ and moving in a retarding field. (*b*) The potential and (*c*) the potential-energy barrier.

As an illustration of this law, consider two parallel plates *A* and *B* separated by a distance *d*, as shown in Fig. 1-1*a*, and with *B* at a negative potential $V_d$ with respect to *A*. An electron leaves the surface of *A* toward *B* with an initial velocity $v_0$ in the *x* direction. What speed will the electron have if it reaches *B*?

From the definition Eq. (1-5), it is clear that only differences of potential have meaning; hence, let us arbitrarily ground *A*, that is, consider it to be at zero potential. Then the potential at *B* is $V = -V_d$, and the potential energy is $U = -qV_d$. If we equate the total energy at *A* to that at *B*, we obtain

$$W = \tfrac{1}{2}mv_0^2 = \tfrac{1}{2}mv^2 + qV_d \qquad \text{J} \tag{1-9}$$

This equation indicates that $v$ must be less than $v_0$, which is obviously correct since the electron is moving in a repelling field. The final velocity attained by the electron in this conservative system is independent of the form of the variation of the field distribution between the plates and depends only on the magnitude of the potential difference $V_d$. Note that, if the electron is to reach electrode *B*, its initial velocity must be large enough so that $\tfrac{1}{2}mv_0^2 > qV_d$. Otherwise, Eq. (1-9) leads to the impossible result that $v$ is imaginary. We now wish to elaborate on these considerations.

**The Concept of a Potential-Energy Barrier** For the configuration of Fig. 1-1*a* in which the electrodes are large in comparison with the separation *d*, we can draw (Fig. 1-1*b*) a linear plot of potential *V* versus distance *x* (in the interelectrode space). The corresponding potential energy *U* versus *x* is indicated in Fig. 1-1*c* and is obtained from the curve by multiplying each ordinate by the charge on the electron (a negative number). The total energy *W* of the electron remains constant and is represented as a horizontal line. The kinetic energy at any distance $x_1$ equals the difference between the total energy *W* and the potential energy *U* at this point. This difference is greatest at 0, indicating that the kinetic energy

is a maximum when the electron leaves the electrode $A$. At point $P$ this difference is zero, which means that no kinetic energy exists, so that the particle is at rest at this point. This distance $x_0$ is the maximum that the electron can travel from $A$. At point $P$ (where $x = x_0$) it comes momentarily to rest and then reverses its motion and returns to $A$.

Consider a point such as $x_2$ which is at a greater distance than $x_0$ from electrode $A$. Here the total energy $W$ is less than the potential energy $U$, so that the difference, which represents the kinetic energy, is negative. This is an impossible physical condition, however, since negative kinetic energy ($\frac{1}{2}mv^2 < 0$) implies an imaginary velocity. We must conclude that the particle can never advance a distance greater than $x_0$ from electrode $A$. At point $P$ this difference is zero, which means that no kinetic energy exists, so that the particle is at rest at this point. This distance $x_0$ is the maximum that the electron can travel. The foregoing analysis leads to the very important conclusion that the shaded portion of Fig. 1-1$c$ can never be penetrated by the electron. Thus, at point $P$, the particle acts as if it had collided with a solid wall, hill, or barrier and the direction of its flight had been altered. Potential-energy barriers of this sort play an important role in the analysis of semiconductor devices.

It must be emphasized that the words "collides with a potential hill" constitute a convenient descriptive phrase and that an actual encounter between two material bodies is not implied.

## 1-2 CONDUCTION IN METALS

In a metal the outer, or valence, electrons of an atom are as much associated with one ion as with another, so that the electron attachment to any individual atom is almost zero. Depending on the metal, at least one (and sometimes two or three) electron (or electrons) per atom is (are) free to move throughout the interior of the metal under the action of applied fields.

Figure 1-2 is a two-dimensional schematic diagram of the charge distribution within a metal. The shaded regions represent the net positive charge of the

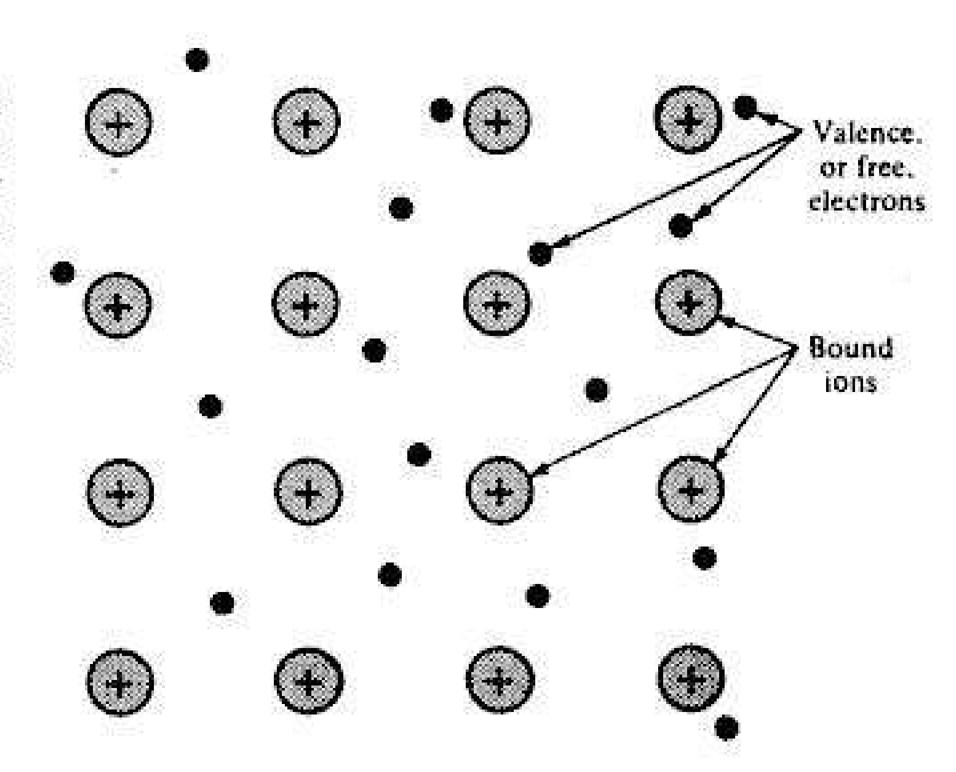



**FIGURE 1-2** Schematic arrangement of the atoms in one plane in a metal, drawn for monovalent atoms. The black dots represent the electron gas, each atom having contributed one electron to this gas.

nucleus and the tightly bound inner electrons. The black dots represent the outer, or valence, electrons in the atom. It is these electrons that cannot be said to belong to any particular atom; instead, they have completely lost their individuality and can wander freely about from atom to atom in the metal. Thus a metal is visualized as a region containing a periodic three-dimensional array of heavy, tightly bound ions permeated with a swarm of electrons that may move about quite freely. This scheme is known as the *electron-gas* description of a metal.

According to the electron-gas theory of a metal, the electrons are in continuous motion, the direction of flight being changed at each collision with the heavy, almost stationary ions. The average distance between collisions is called the *mean free path*. Since the motion is random, the net number of electrons crossing a unit area in the metal during a given time is, on the average, zero. Consequently, the average current is zero.

Let us now see how the situation is changed if a constant electric field $\mathscr{E}$ is applied to the metal. As a result of this electrostatic force, the electrons would be accelerated and the velocity would increase indefinitely with time, were it not for the collisions with the ions. At each inelastic collision with an ion, however, an electron loses energy and changes direction. The probability that an electron moves in a particular direction after a collision is equal to the probability that it travels in the opposite direction. Hence the velocity of an electron increases linearly with time between collisions and is, on an average, reduced to zero at each collision. A steady-state condition is reached when an average value of *drift velocity* $v_d$ is attained; its direction is opposite to that of the electric field. The speed at time $t$ between collisions is $at$, where the acceleration $a$ equals $\mathscr{E}\, q/m$. As a result, the drift velocity is proportional to $\mathscr{E}$ and is given by

$$v_d = \mu\mathscr{E} \qquad \text{m/s} \tag{1-10}$$

where the proportionality constant $\mu$ is called the *mobility* of the electron.[1]

According to the foregoing theory, a steady-state drift velocity has been superimposed on the random thermal motion of the electrons. Such a directed flow of electrons constitutes a current which we can now calculate.

## Current Density

In Fig. 1-3, $N$ electrons are distributed uniformly throughout a conductor of length $L$ and cross-sectional area $A$. An electron, under the influence of an

[1]A subscript is usually added to $\mu$ where more than one type of charge carrier is present. The dimensions of mobility are given in square meters per volt-second [$\text{m}^2/(\text{V}\cdot\text{s})$].

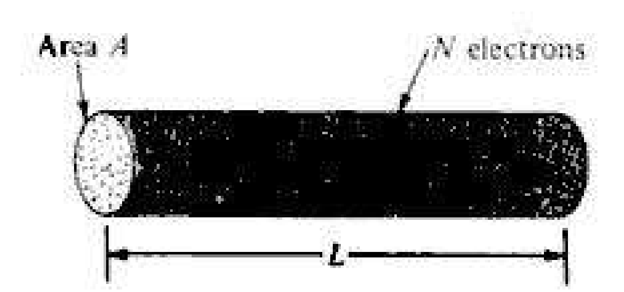



**FIGURE 1-3**
Conductor used to calculate current density.

electric field $\mathscr{E}$ travels $L$ meters in $T$ seconds, thus making the drift velocity $v_d$ equal to $L/T$. The current $I$ is, by definition, the total charge passing through any area per unit time and is the charge per carrier multiplied by the number of carriers per second crossing the area. Hence [in amperes (A)]

$$I = \frac{qN}{T} \cdot \frac{L}{L} = \frac{qNv_d}{L} \qquad \text{A} \tag{1-11}$$

The current density, denoted by the symbol $J$, is the current per unit area in the conducting medium; that is, assuming a uniform current density

$$J \equiv \frac{I}{A} \qquad \text{A/m}^2 \tag{1-12}$$

substitution of Eq. (1-11) into Eq. (1-12) yields

$$J = \frac{qNv_d}{LA} \qquad \text{A/m}^2 \tag{1-13}$$

From Fig. 1-3 it is evident that $LA$ is the volume containing the $N$ electrons. The volume concentration of electrons or simply the electron concentration $n$ is then

$$n = \frac{N}{LA} \qquad \text{m}^{-3} \tag{1-14}$$

and Eq. (1-13) reduces to

$$J = qnv_d = \rho_v \cdot v_d \qquad \text{A/m}^2 \tag{1-15}$$

where $\rho_v = qn$ is the volume charge density in coulombs per cubic meter (C/m$^3$).

This derivation is independent of the form of the conduction medium. Consequently, Fig. 1-3 does not necessarily represent a metal conductor. It may represent equally well a portion of a gaseous-discharge tube or a volume element of a semiconductor. Furthermore, neither $\rho_v$ nor $v_d$ must be constant but may vary with time.

**Conductivity**

We have already seen that $v_d$ is proportional to $\mathscr{E}$. Combination of Eqs. (1-10) and (1-15) results in

$$J = qnv_d = qn\mu\mathscr{E} = \sigma\mathscr{E} \qquad \text{A/m}^2 \tag{1-16}$$

where

$$\sigma = qn\mu \; (\Omega\cdot\text{m})^{-1} \tag{1-17}$$

is the conductivity of the material. By recalling that $\mathscr{E}L = V$ is the applied voltage across the conductor, we can obtain the current $I$ from Eq. (1-16) and recognize it as Ohm's law. Thus

$$I = JA = \sigma\mathscr{E}A \cdot \frac{L}{L} = \frac{\sigma A}{L} V = \frac{V}{R} \qquad \text{A} \tag{1-18}$$

The resistance $R$ of the conductor is given [in ohms ($\Omega$)] by

$$R = \frac{L}{\sigma A} = \rho \frac{L}{A} \quad \Omega \tag{1-19}$$

where the resistivity $\rho$ is the reciprocal of the conductivity.

As already mentioned, the energy acquired by the electrons from the applied field is, as a result of collisions, given to the lattice ions. Consequently, power is dissipated within the metal and the power density [Joule heat in watts per cubic meter (W/m$^3$)] is given by $J = \sigma \mathcal{E}^2$. (This relation is analogous to $P = VI = V^2/R$.)

***Example 1-1*** A conducting line on an IC chip is 2.8 millimeters (mm) long and has a rectangular cross section $1 \times 4$ micrometers ($\mu$m). A current of 5 mA produces a voltage drop of 100 mV across the line. Determine the electron concentration given that the electron mobility is 500 cm$^2$/(V·s).

***Solution*** The electron concentration can be obtained from $\sigma$ in Eq. (1-17). The conductivity is determined by solving Eq. (1-18) for $\sigma$ as

$$\sigma = \frac{IL}{VA} = \frac{5\times10^{-3} \times 2.8\times10^{-3}}{0.1 \times (10^{-6} \times 4\times10^{-6})} = 3.50 \times 10^7 \ (\Omega\text{·m})^{-1}$$

Then, from Eq. (1-17), we obtain

$$n = \frac{\sigma}{q\mu} = \frac{3.5 \times 10^7}{1.60\times10^{-19} \times 500\times10^{-4}}$$
$$= 4.38 \times 10^{27} \text{ m}^{-3} = 4.38 \times 10^{21} \text{ cm}^{-3}$$

As seen in Eq. (1-17), conductivity is proportional to the concentration of charge carriers. The free-electron concentration found in Example 1-1 is a typical value for conductors. Few carriers are available in insulators, and electron concentrations are in the order of $10^7$ m$^{-3}$. Materials whose carrier concentrations lie between those of conductors and insulators are called *semiconductors*, the properties of which are discussed in the next two sections.

## 1-3 THE INTRINSIC SEMICONDUCTOR

Silicon, germanium, and gallium arsenide are the three most widely used semiconductors. Because of the predominance of silicon devices, we confine our discussion to it.

The crystal structure of silicon consists of a regular repetition in three dimensions of a unit cell having the form of a tetrahedron with an atom at each vertex. A two-dimensional symbolic representation of this structure is illustrated in Fig. 1-4. Silicon has a total of 14 electrons in its atomic structure, four of which are valence electrons, so that the atom is tetravalent. The inert ionic core of the silicon atom has a charge of +4 measured in units of electronic charge. The binding forces between neighboring atoms result from the fact that

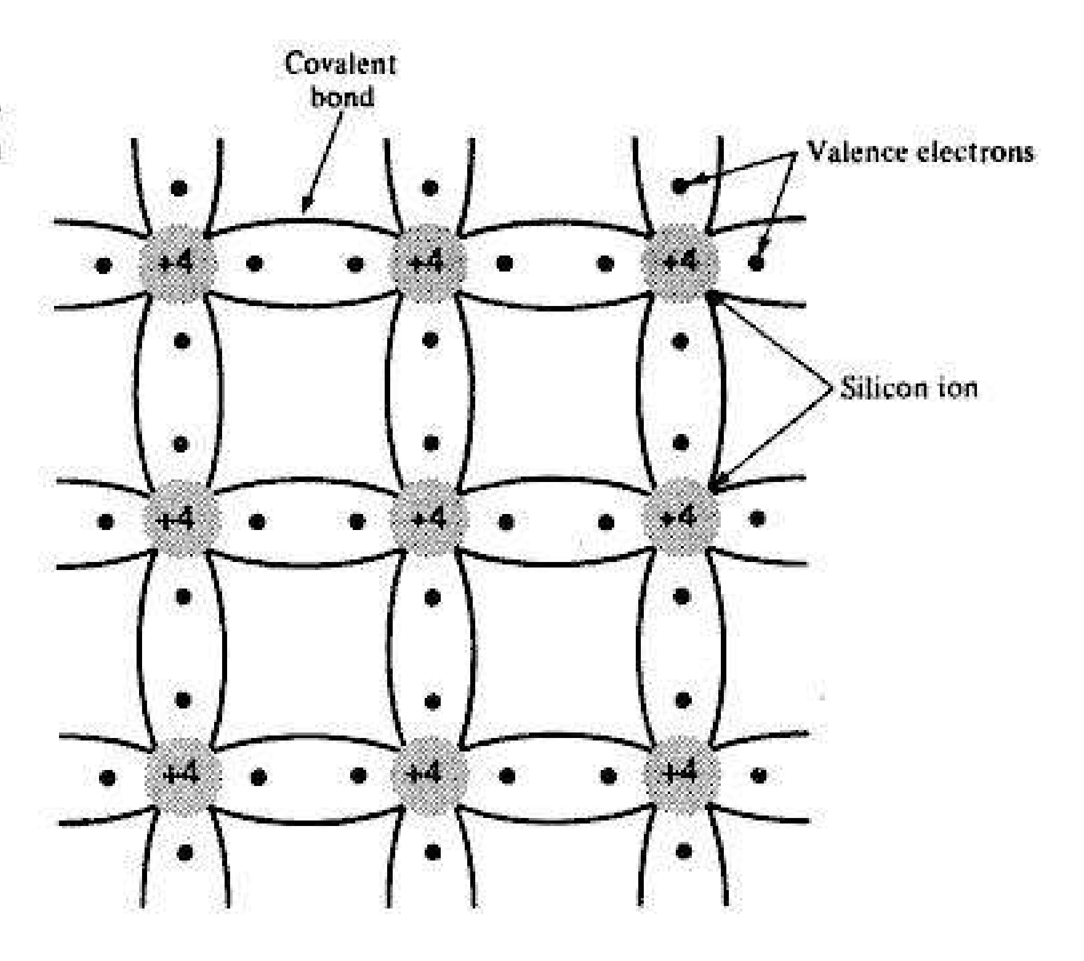



**FIGURE 1-4** Two-dimensional representation of silicon crystal.

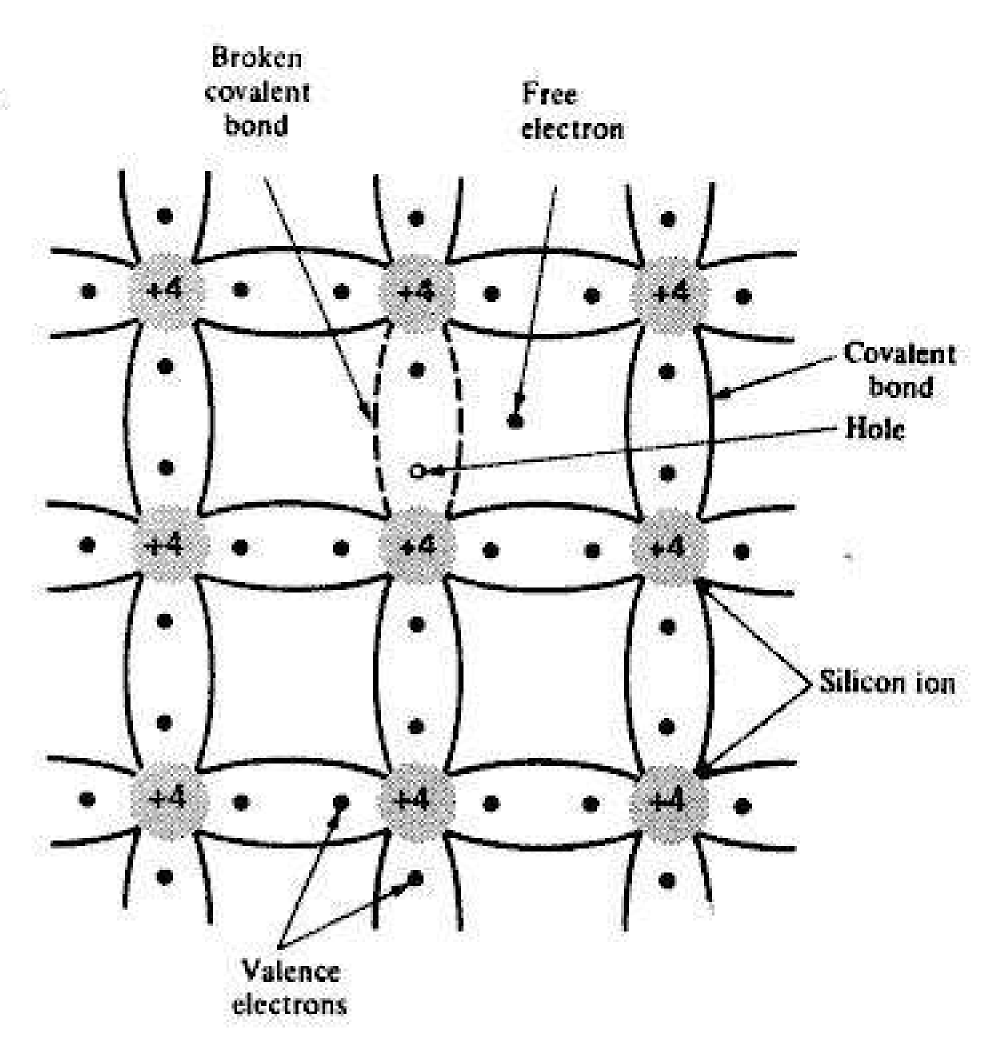
each valence electron of a silicon atom is shared by one of its four nearest neighbors. This covalent bond is represented in Fig. 1-4 by the two lines which join each ion to each of its neighbors. The valence electrons serve to bind one atom to the next and this results in these electrons being tightly bound to the nucleus. Hence, in spite of the availability of four valence electrons, few of these are free to contribute to conduction.

## The Hole

At a very low temperature (say, 0 K) the ideal structure shown in Fig. 1-4 is approached, and the crystal behaves as an insulator, since no free carriers of electricity are available. At room temperature, however, some of the covalent bonds will be broken because of the thermal energy supplied to the crystal, and conduction is made possible. This situation is illustrated in Fig. 1-5. Here an electron, which usually forms part of a covalent bond, is pictured as being dislodged and is thus free to wander in a random fashion throughout the crystal. The energy $E_G$ required to break such a covalent bond is about 1.1 eV for silicon at room temperature. The absence of the electron in the covalent bond is represented by the small circle in Fig. 1-5, and such an incomplete covalent bond is called a *hole*. The importance of the hole is that it may serve as a carrier of electricity comparable in effectiveness with the free electron.

The mechanism by which a hole contributes to the conductivity is qualitatively described as follows. When a bond is incomplete so that a hole exists, it is relatively easy for a valence electron in a neighboring atom to leave its covalent bond to fill this hole. An electron moving from a bond to fill a hole leaves a hole in its initial position. Hence the hole effectively moves in the direction opposite to that of the electron. This hole, in its new position, may now be filled by an electron from another covalent bond, and the hole will

**FIGURE 1-5**
Silicon crystal with a broken covalent bond.




correspondingly move one more step in the direction opposite to the motion of the electron. Here we have a mechanism for the conduction of electricity which does not involve free electrons. This phenomenon is illustrated schematically in Fig. 1-6, where a circle with a dot in it represents a completed bond and an empty circle designates a hole. Figure 1-6*a* shows a row of 10 ions, with a broken bond, or hole, at ion 6. Now imagine that an electron from ion 7 moves into the hole at ion 6, so that the configuration seen in Fig. 1-6*b* results. If we compare this figure with Fig. 1-6*a*, it looks as if the hole in Fig. 1-6*a* has moved toward the right in Fig. 1-6*b* (from ion 6 to ion 7). This discussion indicates that the motion of the hole in one direction actually means the transport of a negative charge an equal distance in the opposite direction. As far as the flow of electric current is concerned, the hole behaves like a positive charge equal in magnitude to the electronic charge. We can consider that the holes are physical entities whose movement constitutes a flow of current. The heuristic argument that a hole behaves as a free positive charge carrier may be justified by quantum mechanics.

**Conduction in Intrinsic Semiconductors** The crystal structure displayed in Figs. 1-4 and 1-5 assumed a pure sample of silicon; that is, the sample contains no foreign atoms. Such pure crystals are called *intrinsic semiconductors*. As shown in

**FIGURE 1-6**
The mechanism by which a hole contributes to the conductivity.

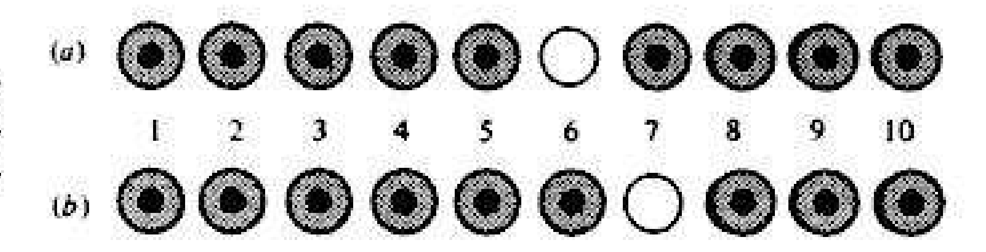


Fig. 1-5, breaking a covalent bond results in both a free electron and a hole. Consequently, the hole concentration $p$ and electron concentration $n$ must be equal and

$$p = n = n_i \tag{1-20}$$

where the intrinsic concentration is denoted by $n_i$. Thermal agitation generates new electron-hole pairs, whereas other electron-hole pairs disappear as a result of recombination. The value of $n_i$ is temperature-dependent; this variation is discussed in Sec. 1-5.

Both holes and electrons contribute to the conduction process. Because the mechanisms by which holes and electrons move about in the crystal differ, the mobilities of these carriers are different. The subscripts $p$ and $n$ are used to distinguish hole and electron values. These carriers move in opposite directions in an electric field, but as they are of opposite sign, the current of each is in the same direction. The current density $J$ that results from an electric field $\mathscr{E}$ is obtained from Eq. (1-16), modified to include both carriers, and is

$$J = q(n\mu_n + p\mu_p)\mathscr{E} = \sigma\mathscr{E} \qquad \text{A/m}^2 \tag{1-21}$$

The conductivity is

$$\sigma = q(n\mu_n + p\mu_p) \qquad (\Omega\cdot\text{m})^{-1} \tag{1-22}$$

For intrinsic semiconductors, $p = n = n_i$ and Eq. (1-22) reduces to

$$\sigma_i = qn_i(\mu_n + \mu_p) \qquad (\Omega\cdot\text{m})^{-1} \tag{1-23}$$

Values of important properties of silicon are given in Table 1-1. Note that silicon has on the order of $10^{22}$ atoms/cm$^3$, while at room temperature (300 K),

**TABLE 1-1 Properties of Intrinsic Silicon**

| *Property* | *Value* |
|---|---|
| Atomic number | 14 |
| Atomic weight | 28.1 |
| Density (g/cm$^3$) | 2.33 |
| Relative permittivity (dielectric constant) | 11.9 |
| Atoms/cm$^3$ | $5.0 \times 10^{22}$ |
| Energy gap $E_{G0}$ at 0 K (eV) | 1.21 |
| Energy gap $E_G$ at 300 K (eV) | 1.12 |
| Resistivity at 300 K ($\Omega$·cm) | $2.30 \times 10^5$ |
| Electron mobility $\mu_n$ at 300 K [cm$^2$/(V·s)] | 1500 |
| Hole mobility $\mu_p$ at 300 K [cm$^2$/(V·s)] | 475 |
| Intrinsic concentration at 300 K (cm$^{-3}$) | $1.45 \times 10^{10}$ |
| Electron diffusion constant $D_n$ at 300 K (cm$^2$/s) | 34 |
| Hole diffusion constant $D_p$ at 300 K (cm$^2$/s) | 13 |

From S. M. Sze (ed.), "VSLI Technology," McGraw-Hill Book Company, New York, 1983.

$n_i \simeq 10^{10}$ cm$^{-3}$. Hence only 1 atom in about $10^{12}$ contributes a free electron (and also a hole) to the crystal because of broken covalent bonds.

*Example 1-2*

An intrinsic silicon bar is 3 mm long and has a rectangular cross section 50 × 100 μm. At 300 K, determine the electric field intensity in the bar and the voltage across the bar when a steady current of 1 μA is measured.

*Solution*

The field intensity can be obtained from the current density and conductivity as

$$\mathscr{E} = \frac{J}{\sigma} = \frac{I}{A} \times \frac{1}{\sigma} = \frac{I}{A} \cdot \rho \qquad \text{V/m}$$

Using the value of $\rho$ given in Table 1-1, we obtain

$$\mathscr{E} = \frac{10^{-6}}{50 \times 10^{-6} \times 100 \times 10^{-6}} \times 2.30 \times 10^{5} \times 10^{-2}$$

where the factor $10^{-2}$ converts the resistivity from Ω·cm to Ω·m:

$$\mathscr{E} = 4.60 \times 10^{5} \text{ V/m} = 4.60 \times 10^{3} \text{ V/cm}$$

The voltage across the bar is

$$V_{\text{bar}} = \mathscr{E}L = 4.60 \times 10^{5} \times 3 \times 10^{-3} = 1380 \text{ V}$$

The result obtained in Example 1-2 indicates that an extremely large voltage is needed to produce a small current (1 μA). This, however, is not surprising since the intrinsic carrier concentration is much closer to that of an insulator than it is to a conductor. Thus intrinsic semiconductors are not suitable for electron devices. In Sec. 1-4 we investigate one method by which carrier concentration can be increased.

## 1-4 EXTRINSIC SEMICONDUCTORS

A common expedient used to increase the number of carriers is to introduce a small carefully controlled impurity content into an intrinsic semiconductor. The addition of impurities, most often trivalent or pentavalent atoms, forms an *extrinsic*, or *doped*, semiconductor. Each type of impurity establishes a semiconductor which has a predominance of one kind of carrier. The usual level of doping is in the range of 1 impurity atom for $10^6$ to $10^8$ silicon atoms. Thus most physical and chemical properties are essentially those of silicon and only the electrical properties change markedly.

**n-Type Semiconductors** Figure 1-7 depicts the crystal structure obtained when silicon is doped with a pentavalent impurity. Four of the five valence electrons occupy covalent bonds, and the fifth will be nominally unbound and will be available as a carrier of current. The energy required to detach this fifth electron from the atom is of the order of only 0.05 eV for silicon and is considerably less than the energy required to break a covalent bond. Suitable pentavalent impurities are anti-

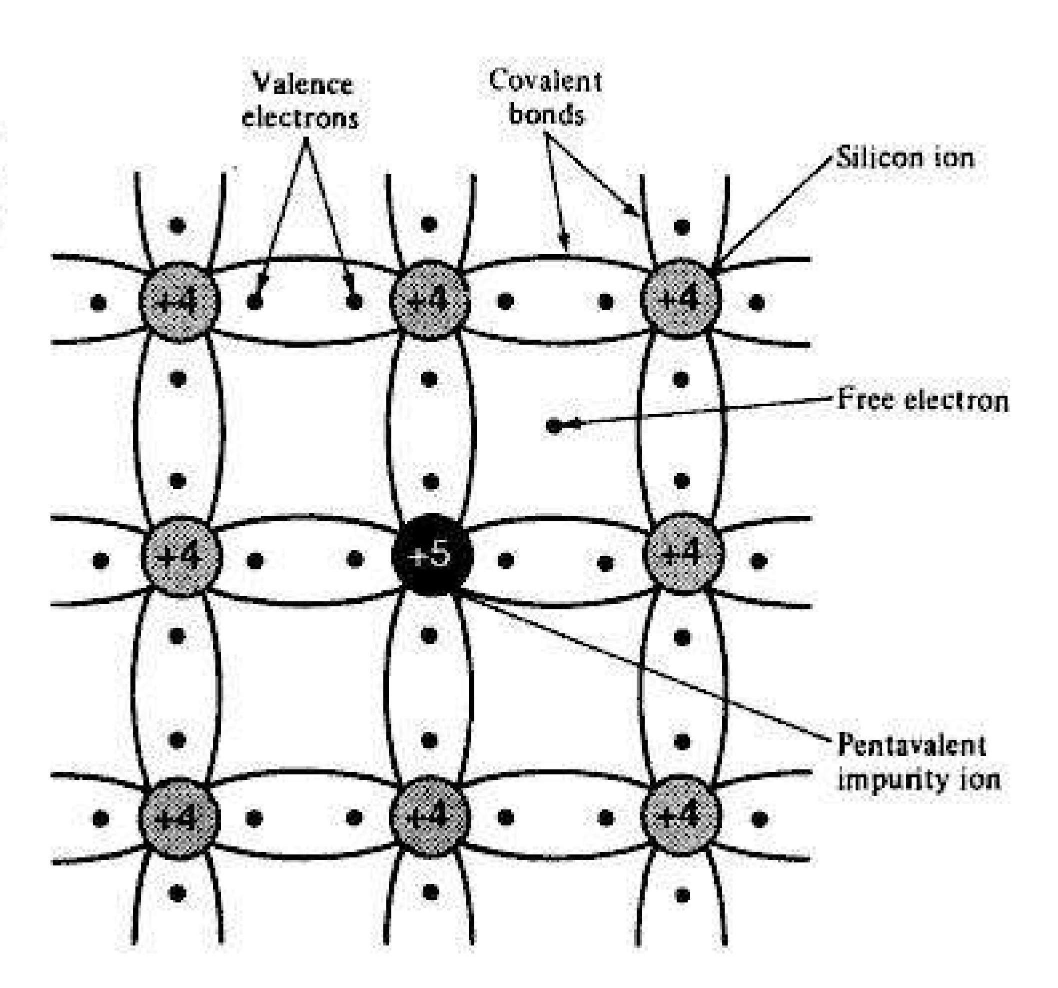



**FIGURE 1-7** Crystal lattice with a silicon atom displaced by a pentavalent impurity atom.

mony, phosphorus, and arsenic. Such impurities donate excess electron carriers and are referred to as *donor*, or *n-type*, impurities.

If intrinsic semiconductor material is doped with $n$-type impurities, not only does the number of electrons increase, but the number of holes decreases below that which would be available in the intrinsic semiconductor. The number of holes decreases because the larger number of electrons present causes the rate of recombination of electrons with holes to increase. Consequently, the dominant carriers are the negative electrons and doping with donors results in an *n-type semiconductor*.

**p-Type Semiconductors** Boron, gallium, and indium are trivalent atoms which, when added to intrinsic semiconductors, provide electrons to fill only three covalent bonds. The vacancy that exists in the fourth bond constitutes a hole as illustrated in Fig. 1-8. This type of impurity makes positive carriers available because it creates holes which can accept electrons. Thus trivalent impurities are called *acceptors* and form *p-type semiconductors* in which holes are the predominant carrier.

**The Mass-Action Law** We noted previously that the addition of $n$-type impurities causes the number of holes to decrease. Similarly, doping with $p$-type impurities decreases the concentration of free electrons below that in the intrinsic semiconductor. A theoretical analysis (Sec. 1-7) leads to the result that, under thermal equilibrium, the product of the free negative and positive concentrations is a constant independent of the amount of donor and acceptor impurity doping. This relationship is called the *mass-action law* and is given by

$$np = n_i^2 \tag{1-24}$$

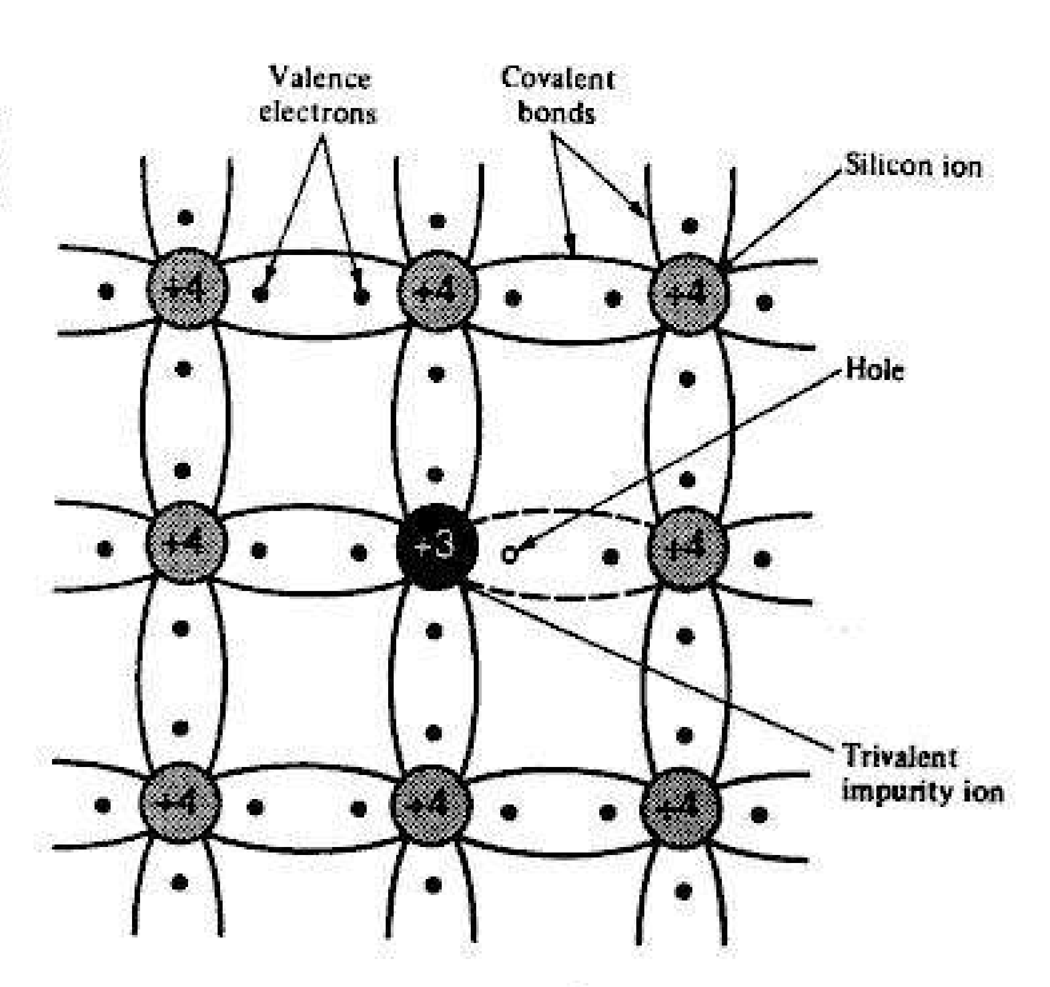



**FIGURE 1-8**
Crystal lattice with a silicon atom displaced by a trivalent impurity atom.

The intrinsic concentration $n_i$ is a function of temperature (Sec. 1-5).

We have the important result that the doping of an intrinsic semiconductor not only increases the conductivity, but also serves to produce a conductor in which the charge carriers are either predominantly holes or predominantly electrons. In an $n$-type semiconductor, the electrons are called the *majority carriers*, and the holes are called the *minority carriers*. In a $p$-type material, the holes are the majority carriers and the electrons are the minority carriers.

**Carrier Concentrations** We have previously indicated that only a small amount of energy is needed to ionize the impurity atoms. The temperatures at which electronic devices normally operate (>200 K) provide sufficient thermal energy to ionize virtually all impurities. This fact, in conjunction with the mass-action law, permits us to determine the charge densities in a semiconductor.

Let $N_D$ be the concentration of donor atoms and $N_A$ the concentration of acceptor atoms. Since these impurities are practically all ionized, they produce positive-ion and negative-ion densities of $N_D$ and $N_A$, respectively. To maintain the electric neutrality of the crystal, the total positive charge density must equal the concentration of negative charges as expressed in Eq. (1-25):

$$N_D + p = N_A + n \tag{1-25}$$

Let us now consider an $n$-type material having $N_A = 0$. Since the number of electrons is much greater than the number of holes in an $n$-type semiconductor ($n \gg p$), Eq. (1-25) reduces to

$$n \simeq N_D \tag{1-26}$$

*In an n-type material the free-electron concentration is approximately equal to the density of donor atoms.*

The concentration $p$ of holes in the $n$-type semiconductor is obtained from Eq. (1-24). Thus

$$p = \frac{n_i^2}{N_D} \tag{1-27}$$

Similarly, in a $p$-type semiconductor, with $N_D = 0$, we have

$$p \simeq N_A \tag{1-28}$$

and

$$n = \frac{n_i^2}{N_A} \tag{1-29}$$

***Example 1-3*** An $n$-type silicon sample is 3 mm long and has a rectangular cross section $50 \times 100\ \mu\text{m}$. The donor concentration at 300 K is $5 \times 10^{14}\ \text{cm}^{-3}$ and corresponds to 1 impurity atom for $10^8$ silicon atoms. A steady current of $1\ \mu\text{A}$ exists in the bar. Determine the electron and hole concentrations, the conductivity, and the voltage across the bar. (Note that this is an $n$-type sample that has the same dimensions and current as does the intrinsic silicon in Example 1-2.)

***Solution***[1] From Eqs. (1-26) and (1-27), and using the values of $n_i$ and $\mu_n$ in Table 1-1, we obtain

$$n = N_D = 5 \times 10^{14}\ \text{cm}^{-3}$$

and

$$p = \frac{(1.45 \times 10^{10})^2}{5 \times 10^{14}} = 4.2 \times 10^5\ \text{cm}^{-3}$$

As $n \gg p$, only the electron concentration need be considered in Eq. (1-22), so that the conductivity is

$$\sigma = qn\mu_n = 1.60 \times 10^{-19} \times 5 \times 10^{14} \times 1.5 \times 10^3 = 0.12\ (\Omega\text{·cm})^{-1}$$

The voltage across the bar, obtained from $\mathcal{E} \cdot L$, where, by means of Eq. (1-21), $\mathcal{E} = J/\sigma$, is

$$V_{\text{bar}} = \frac{J}{\sigma} L = \frac{IL}{A\sigma} = \frac{10^{-6}\,(0.3)}{(5 \times 10^{-3})\,(10^{-2}) \times 0.12} = 0.05\ \text{V}$$

The efficacy of using extrinsic semiconductors in electronic devices is readily apparent when the results of Examples 1-2 and 1-3 are compared. To produce the small current of $1\ \mu\text{A}$, 1380 V must be applied to the intrinsic sample,

---

[1]All calculations in this example are made by using centimeters instead of meters in all equations.

whereas only 50 mV is required for the $n$-type sample. This reduction of voltage by a factor of 28,000 exactly equals the decrease in resistivity (from $2.30 \times 10^5$ to $1/\sigma = 8.33\ \Omega\cdot$cm). Yet the dramatic increase in the number of free electrons ($1.45 \times 10^{10}$ to $5 \times 10^{14}$ cm$^{-3}$) occurs when only 1 silicon atom in 100 million is replaced by an impurity atom!

It is possible to add donors to a $p$-type crystal or, conversely, to add acceptors to $n$-type material. If equal concentrations of donors and acceptors permeate the semiconductor, the semiconductor remains intrinsic. The hole of the acceptor combines with the conduction electron of the donor to give no additional free carriers. Thus, from Eq. (1-25) with $N_D = N_A$, we observe that $p = n$, and from Eq. (1-24), $n^2 = n_i^2$, or $n = n_i$ = intrinsic concentration.

An extension of the preceding argument indicates that if the concentration of donor atoms added to a $p$-type semiconductor exceeds the acceptor concentration ($N_D > N_A$), the specimen is changed from a $p$-type to an $n$-type semiconductor. Conversely, the addition of a sufficient number of acceptors to an $n$-type sample results in a $p$-type semiconductor. This is precisely what is done in fabricating IC transistors. To determine the carrier concentrations under these circumstances, $N_D$ is replaced by $N_D - N_A$ in Eqs. (1-26) and (1-27) when $p$-type material is changed to an $n$-type semiconductor. Similarly, when an $n$-type semiconductor is converted to $p$ type, $N_A$ in Eqs. (1-28) and (1-29) is replaced by $N_A - N_D$.

**Generation and Recombination of Charges** In an intrinsic semiconductor the number of holes is equal to the number of free electrons. Thermal agitation, however, continues to generate $g$ new hole-electron pairs per unit volume per second, while other hole-electron pairs disappear as a result of recombination; in other words, free electrons fall into empty covalent bonds, resulting in the loss of a pair of mobile carriers. On an average, a hole (an electron) will exist for $\tau_p$ ($\tau_n$) seconds before recombination. This time is called the *mean lifetime* of the hole (electron). These parameters are very important in semiconductor devices because they indicate the time required for electron and hole concentrations which have been caused to change to return to their equilibrium concentrations.

## 1-5 VARIATIONS IN THE PROPERTIES OF SILICON

The conductivity of a semiconductor, given in Eq. (1-22), depends on both hole and electron concentration and mobility. Because semiconductor devices are subject to a wide range of operating temperatures, the variations of these parameters with temperature are important.

**Intrinsic Concentration** With increasing temperature, the density of hole-electron pairs increases in an intrinsic semiconductor. Theoretically it is found that the intrinsic concentration $n_i$ varies with $T$ as

$$n_i^2 = A_0 T^3 \epsilon^{-E_{GO}/kT} \tag{1-30}$$

where $E_{G0}$ is the energy gap (the energy required to break a covalent bond) at 0 K in electron volts, $k$ is the Boltzmann constant in electron volts per degree kelvin (eV/K), and $A_0$ is a constant independent of $T$.

The increase in $n_i^2$ with temperature also has an effect on the charge densities in extrinsic semiconductors. For example, consider an $n$-type sample with a donor concentration $N_D$ which is subjected to a temperature increase from 300 to 400 K. The electron density $n$ at 400 K does not change appreciably from its value at 300 K because the ionized donor impurities contribute nearly all of the carriers. However, the mass-action law indicates that the hole concentration $p$ increases. Similarly, for $p$-type semiconductors, $n$ increases with modest rises in temperature, and $p \simeq N_A$ remains constant.

**Mobility**

The variation with temperature (100 to 400 K) of electron and hole mobilities is proportional to $T^{-m}$. For silicon, $m = 2.5$ for electrons and 2.7 for holes. Mobility $\mu$ decreases with temperature because more carriers are present and these carriers are more energetic at higher temperatures. Each of these facts results in an increased number of collisions and $\mu$ decreases.

Mobilities are also functions of the electric field intensity and doping levels. In $n$-type silicon, $\mu$ is constant at a given temperature only if $\mathscr{E} < 10^3$ V/cm. For $\mathscr{E} > 10^4$ V/cm, $\mu_n$ is inversely proportional to $\mathscr{E}$ and drift velocities approach $10^7$ cm/s (the saturation velocity). Between $10^3$ and $10^4$ V/cm, $\mu_n$ varies approximately as $\mathscr{E}^{-1/2}$.

**Conductivity**

The conductivity of an intrinsic semiconductor increases with increasing temperature because the increase in hole-electron pairs is greater than the decrease in their mobilities. For extrinsic semiconductors, in the temperature range 100 to 600 K, the number of majority carriers is nearly constant but diminished mobility causes the conductivity to decrease with temperature.

## 1-6 DIFFUSION

In addition to a conduction current, the transport of charges in a semiconductor may be accounted for by a mechanism called *diffusion* (not ordinarily encountered in metals). The essential features of diffusion are now discussed.

It is possible to have a nonuniform concentration of particles in a semiconductor. As indicated in Fig. 1-9, the concentration $p$ of holes varies with distance

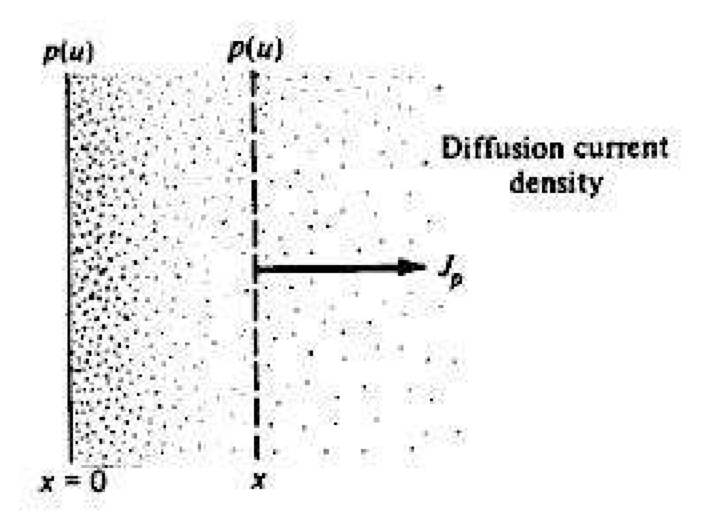



**FIGURE 1-9** Representation of a nonuniform hole density and the resultant diffusion current density.

$x$ in the semiconductor, and there exists a concentration gradient $dp/dx$ in the density of carriers. The existence of a gradient implies that if an imaginary surface (indicated by dashed line) is drawn in the semiconductor, the density of holes immediately on one side of the surface is larger than the density on the other side. The holes are in a random motion as a result of their thermal energy. Accordingly, holes will continue to move back and forth across this surface. We may then expect that, in a given time interval, more holes will cross the surface from the side of greater concentration to the side of smaller concentration than in the reverse direction. This net transport of holes across the surface constitutes a current in the positive $x$ direction. It should be noted that this net transport of charge is not the result of mutual repulsion among charges of like sign, but is simply the result of a statistical phenomenon. This diffusion is exactly analogous to that which occurs in a contained neutral gas if a concentration gradient exists.[1] The diffusion hole-current density $J_p$ is proportional to the concentration gradient and is given by

$$J_p = -qD_p \frac{dp}{dx} \qquad \text{A/m}^2 \tag{1-31}$$

where $D_p$ (in square meters per second) is called the *diffusion constant* for holes. Since $p$ in Fig. 1-9 decreases with increasing $x$, $dp/dx$ is negative and the minus sign in Eq. (1-31) is needed, so that $J_p$ will be positive in the positive $x$ direction. A similar equation exists for diffusion electron current density [$p$ is replaced by $n$, and the minus sign is replaced by a plus sign in Eq. (1-31)].

**The Einstein Relationship** Since both diffusion and mobility are statistical thermodynamic phenomena, $D$ and $\mu$ are not independent. The relationship between them is given by the Einstein equation

$$\frac{D_p}{\mu_p} = \frac{D_n}{\mu_n} = V_T \tag{1-32}$$

where $V_T$ is the "volt equivalent of temperature," defined by

$$V_T \equiv \frac{\bar{k}T}{q} = \frac{T}{11{,}600} \qquad \text{V} \tag{1-33}$$

where $\bar{k}$ is the Boltzmann constant in joules per Kelvin. Note the distinction between $\bar{k}$ and $k$; the latter is the Boltzmann constant in electron volts per Kelvin. (Values of $\bar{k}$ and $k$ are given in App. A-1. From Sec. 1-3 it follows that $\bar{k} = 1.60 \times 10^{-19}k$.) At room temperature (300 K), $V_T = 0.0259$ V, and $\mu = 38.6D$. (Measured values of $\mu$ and computed values of $D$ for silicon are given in Table 1-1.)

[1] It is also the process by which the scent of flowers can permeate a room.

**Total Current** It is possible for both a potential gradient and a concentration gradient to exist simultaneously within a semiconductor. In such a situation the total hole current is the sum of the drift current [Eq. (1-16), with $n$ replaced by $p$] and the diffusion current [Eq. (1-31)], or

$$J_p = q\mu_p p\mathscr{E} - qD_p \frac{dp}{dx} \qquad \text{A/m}^2 \tag{1-34}$$

Similarly, the net electron current is

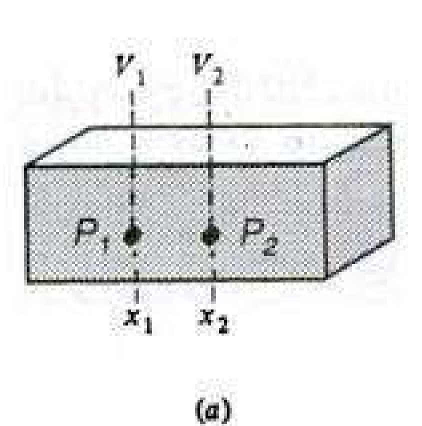
$$J_n = q\mu_n n\mathscr{E} + qD_n \frac{dn}{dx} \qquad \text{A/m}^2 \tag{1-35}$$

## 1-7 GRADED SEMICONDUCTORS

The semiconductor sample shown in Fig. 1-10*a* has a hole concentration that is a function of $x$; that is, the doping is graded (nonuniform). The electron density must also vary with $x$, a consequence of the mass-action law. Let us assume thermal equilibrium and that no carriers are injected into the sample from any *external* source (zero excitation). Under these conditions there can be no *steady* charge motion, only the random motion due to thermal agitation. Hence the total hole current must be zero; also, the total electron current is zero. Since $p$ is not constant, we expect a nonzero hole diffusion current. For the total hole current to vanish, a hole drift current must exist which is equal and opposite to the diffusion current. However, a conduction current requires an electric field, and we conclude that, as a result of the nonuniform doping, an electric field is generated within the semiconductor. We shall now find this field and the corresponding potential variation throughout the bar.

Setting $J_p = 0$ in Eq. (1-34) and using the Einstein relationship $D_p = \mu_p V_T$ [Eq. (1-32)], we obtain

$$\mathscr{E} = \frac{V_T}{p}\frac{dp}{dx} \qquad \text{V/m} \tag{1-36}$$



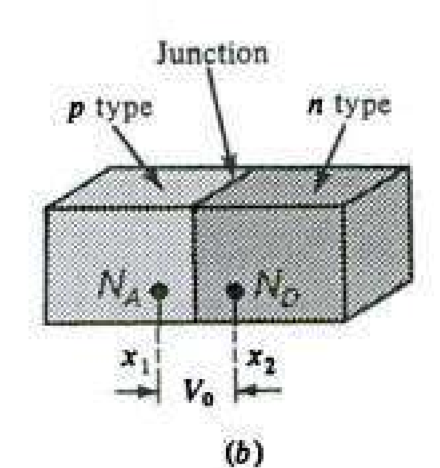




**FIGURE 1-10** (*a*) A graded semiconductor; $p(x)$ is not constant. (*b*) A *pn* junction in which $p$ and $n$ are uniformly doped with impurity concentrations $N_A$ and $N_D$, respectively.

If the doping concentration $p(x)$ is known, this equation allows the built-in field $\mathscr{E}(x)$ to be calculated. From $\mathscr{E} = -dV/dx$, we can calculate the potential as

$$dV = -V_T \frac{dp}{p} \tag{1-37}$$

Integration of Eq. (1-36) from $x_1$, where the concentration is $p_1$ and the potential is $V_1$, to $x_2$ where $p = p_2$ and $V = V_2$, yields

$$V_{21} \equiv V_2 - V_1 = V_T \ln \frac{p_1}{p_2} \qquad \text{V} \tag{1-38}$$

Note that the potential difference between two points depends only on the concentration at these two points and is independent of their separation $x_2 - x_1$. Equation (1-38) may be expressed in the form

$$p_1 = p_2 \epsilon^{+V_{21}/V_T} \tag{1-39}$$

This is the Boltzmann relationship of kinetic gas theory.

**Mass-Action Law** Starting with $J_n = 0$ and proceeding as above, we obtain the Boltzmann equation for electrons:

$$n_1 = n_2 \epsilon^{-V_{21}/V_T} \tag{1-40}$$

Multiplication of Eq. (1-39) by Eq. (1-40) gives

$$n_1 p_1 = n_2 p_2 \tag{1-41}$$

This equation states that the product $np$ is a constant independent of $x$, and hence the amount of doping, under thermal equilibrium. For an intrinsic semiconductor, $n = p = n_i$, and $np = n_i^2$, which is the law of mass action introduced in Eq. (1-24).

**An Open-Circuited Step-Graded Junction** Consider the special case indicated in Fig. 1-10*b*. The left half of the bar is $p$ type with a constant concentration $N_A$, whereas the right half is $n$ type with a uniform density $N_D$. The dashed plane is a metallurgical ($pn$) junction separating the two sections with different concentration. This type of doping, where the density changes abruptly from $p$ to $n$ type, is called *step grading*, and the junction is located at the plane where the concentration is zero. As described earlier, theory indicates that there is a built-in potential between these two sections called the *contact difference of potential* $V_0$. Equation (1-38) allows us to calculate $V_0$ as

$$V_0 = V_{21} = V_T \ln \frac{p_{p0}}{p_{n0}} \qquad \text{V} \tag{1-42}$$

because $p_1 = p_{p0}$ = thermal-equilibrium hole concentration in the $p$ side and $p_2 = p_{n0}$ = thermal equilibrium hole concentration in the $n$ side. From Eq.

(1-28), $p_{p0} = N_A$, and from Eq. (1-27) $p_{n0} = n_i^2/N_D$, so that

$$V_0 = V_T \ln \frac{N_A N_D}{n_i^2} \quad \text{V} \tag{1-43}$$

The same expression for $V_0$ is obtained from an analysis corresponding to that given above and based on equating the total electron current $I_n$ to zero (Prob. 1-18). The *pn* junction is studied in detail in Chap. 2.

## REFERENCES


1 Shockley, W.: "Electrons and Holes in Semiconductors," D. Van Nostrand, Princeton, N.J. (reprinted), 1963.

2 Yang, E. S.: "Fundamentals of Semiconductor Devices," McGraw-Hill Book Company, New York, 1978.

3 Sze, S. M.: "Physics of Semiconductor Devices," 2d ed., John Wiley & Sons, New York, 1979.

4 Adler, R. B., A. C. Smith, and R. L. Longini: "Introduction to Semiconductor Physics," vol. 1, SEEC, John Wiley & Sons, New York, 1965.


## REVIEW QUESTIONS

1-1 Define electric field intensity.

1-2 Define potential energy.

1-3 Define an electron volt.

1-4 What is meant by the electron-gas description of a metal?

1-5 Define mobility.

1-6 Define conductivity.

1-7 Why does an intrinsic semiconductor behave as an insulator at 0 K?

1-8 What is a hole? How does it contribute to conduction?

1-9 (*a*) What is the intrinsic concentration of holes?
(*b*) What is the relationship between the density in Rev. 1-9*a* and the electron concentration?

1-10 What is the distinction between intrinsic and extrinsic semiconductors?

1-11 Show a two-dimensional representation of a silicon crystal containing a donor impurity atom.

1-12 Repeat Rev. 1-11 for an acceptor impurity atom.

1-13 What type of semiconductor results when silicon is doped with (*a*) donor and (*b*) acceptor impurities?

1-14 State the mass-action law.

1-15 A semiconductor has donor and acceptor concentrations of $N_D$ and $N_A$, respectively. What relationships must be used to determine the electron $n$ and hole $p$ concentrations?

1-16 Describe recombination.

1-17 Define mean lifetime of a carrier.

1-18 Does the resistance of an extrinsic semiconductor increase or decrease with temperature? Explain briefly.

1-19 Repeat Rev. 1-18 for an intrinsic semiconductor.

1-20 Define the volt equivalent of temperature.

1-21 What condition(s) must exist for diffusion to occur?

1-22 Define the diffusion constant for (*a*) holes, and (*b*) electrons.

1-23 Are diffusion and drift related? If so, how?

1-24 Write an equation for the net electron current in a semiconductor and give the physical significance of each term.

1-25 Define a graded semiconductor.

1-26 Why must an electric field exist in a graded semiconductor?

1-27 On what parameters does the contact difference of potential in an open-circuited step-graded *pn* junction depend?

# Chapter 2
# THE *pn* JUNCTION DIODE

The *pn* junction is the basic building block on which the operation of all semiconductor devices depends. On the basis of the semiconductor properties described in Chap. 1, the behavior of the *pn* junction is developed. Attention is focused on the volt-ampere characteristics and the circuit models useful in representing junction operation. Because the *pn* junction is itself a two-element device (the diode), we also investigate its use as a circuit element.

## 2-1 THE OPEN-CIRCUITED JUNCTION

A *pn* junction is formed (Fig. 2-1) when a single crystal of semiconductor is doped with acceptors on one side and donors on the other. In Fig. 2-1, donor ions are represented by plus signs and the electrons they "donate" are indicated by the small, filled circles. Holes are depicted by unfilled, small circles and acceptor ions by minus signs. It is assumed that the junction in Fig. 2-1 has reached equilibrium conditions and that the semiconductor has uniform cross section.

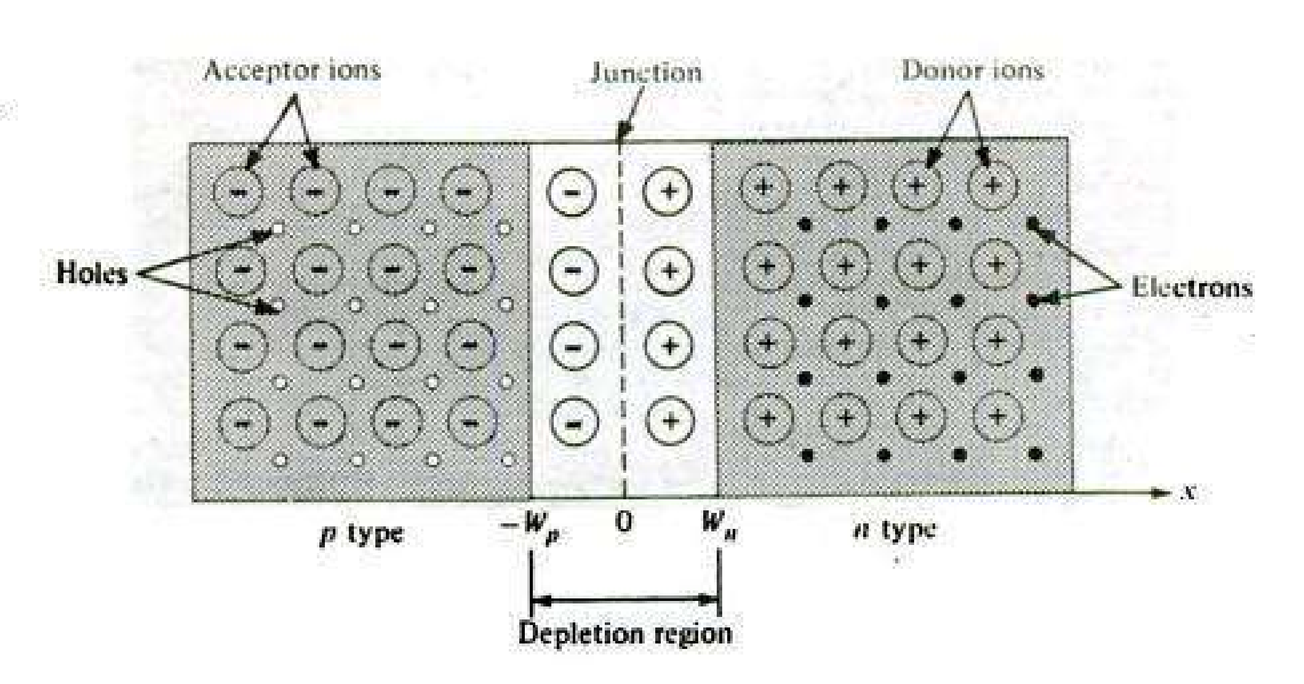



**FIGURE 2-1**
A schematic representation of a *pn* junction.

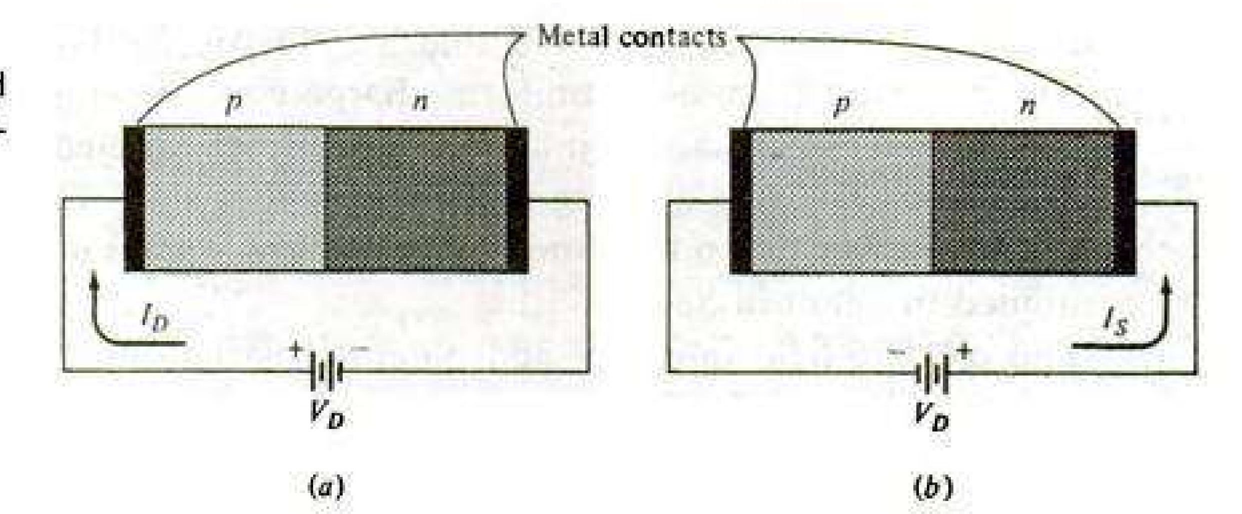



**FIGURE 2-3**
(*a*) The forward-biased and (*b*) the reverse-biased *pn* junction.

**2-2 THE BIASED *pn* JUNCTION** The essential electrical characteristic of a *pn* junction is that carrier flow is permitted in one direction and is virtually eliminated in the other. We now consider how this *unilateral*, or *rectifier*, action comes about by investigating the behavior of the junction to an externally applied voltage.

**The Forward-Biased *pn* Junction** In Fig. 2-3*a* the voltage $V_D$ is applied to the junction with the positive terminal of the battery connected to the *p* side and the negative terminal to the *n* side. In accordance with Fig. 2-2, we assume that no voltage drop appears across the body of the semiconductor outside the depletion region or across the metal contacts. Consequently, the applied voltage reduces the potential barrier by an amount $qV_D$, thus disturbing the equilibrium established between the diffusion and drift of carriers across the junction. The result of decreasing the junction potential is to permit holes to diffuse from the *p* side to the *n* side of the junction. Similarly, electrons can now diffuse from the *n* to the *p* side. Holes traveling to the right and electrons to the left constitute a current in the same direction. Hence the resultant current crossing the junction is the sum of the hole and electron currents. Once they cross the junction, the electrons (holes) become minority carriers in the *p* (*n*) region and constitute an injected minority current. This diffusion current can be large as the number of carriers available is substantial. The applied voltage of the polarity indicated in Fig. 2-3*a* that produces this current is called *forward bias* and the junction is said to be *forward-biased*.

**The Reverse-Biased *pn* Junction** The polarity of the applied voltage in Fig. 2-3*b* (opposite to that shown in Fig. 2-3*a*) *reverse-biases* the junction. The effect of this voltage is to increase the potential barrier by $qV_D$ and, as a result, decreases the flow of majority carriers (holes in *p* type and electrons in *n* type). However, minority carriers (electrons in *p* type and holes in *n* type), since they "fall down" the potential hill, are unaffected by the increase in barrier potential. The initial equilibrium conditions are, nevertheless, perturbed and a small current exists from *n* to *p* across the junction (opposite that forward bias). This current, designated by $I_S$ and called the *reverse saturation current*, is very small as few minority carriers are available. From the preceding discussion we expect $I_S$ to be independent of the magnitude of the reverse bias.

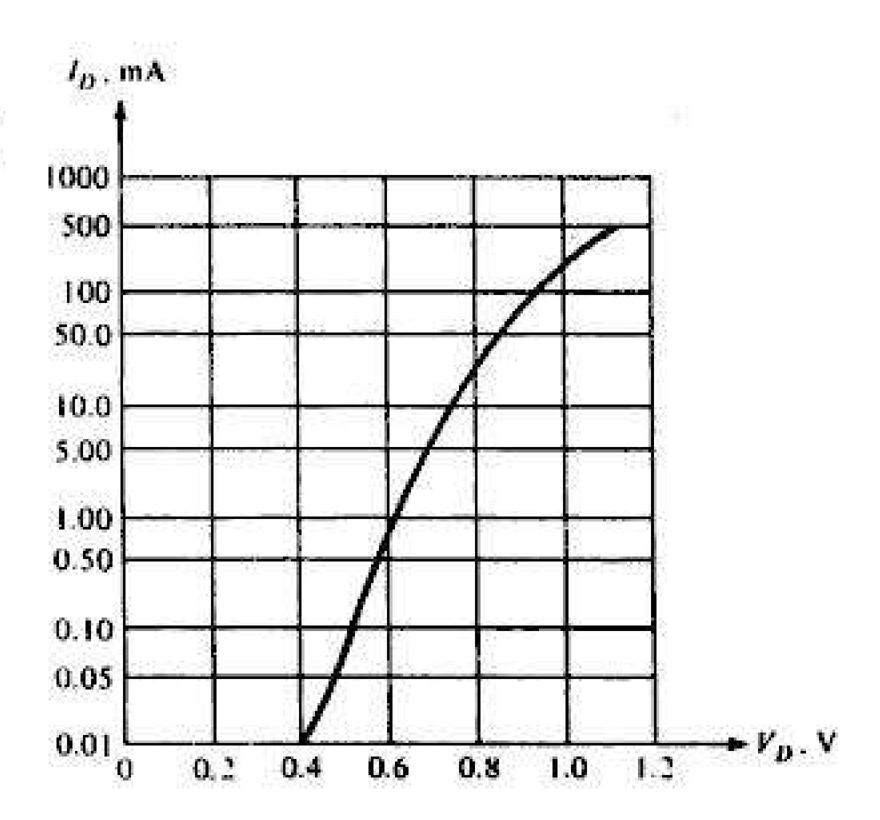



**FIGURE 2-6** Logarithmic characteristic of a 1N4153 silicon diode at 25°C.

A plot of log $I_D$ versus $V_D$ results in a straight line of slope $0.434/\eta V_T$ from which $\eta$ is obtained. In practice, this linear relation is observed at low current levels. For the logarithmic characteristic of the IN4153 at $T = 25°C$, shown in Fig. 2-6, the linear range occurs for $I_D < 25$ mA. From the slope, $\eta$ is approximately 2. At higher current levels, the slope decreases because the total applied voltage does not appear across the junction but also includes the ohmic voltage drop in the contacts and bulk semiconductor. Furthermore, it is found that $\eta$ approaches unity and $I_D = I_S\,\epsilon^{V_D/V_T}$ for high current levels.

***Example 2-1*** Determine the change in diode voltage corresponding to a 10:1 change in $I_D$ for a diode operating at 300 K.

***Solution*** From Eq. (2-5), we obtain

$$I_{D1} = I_S\,\epsilon^{V_{D1}/\eta V_T} \qquad \text{and} \qquad I_{D2} = I_S\,\epsilon^{V_{D2}/\eta V_T}$$

Thus

$$\frac{I_{D2}}{I_{D1}} = \epsilon^{(V_{D2} - V_{D1})/\eta V_T}$$

from which

$$\log\frac{I_{D2}}{I_{D1}} = \frac{0.434(V_{D2} - V_{D1})}{\eta V_T} \qquad \text{and} \qquad V_{D2} - V_{D1} = 2.303\eta V_T \log\frac{I_{D2}}{I_{D1}}$$

At $T = 300$ K, $V_T = 26$ mV from Eq. (2-4) and for $I_{D2}/I_{D1} = 10$, $V_{D2} - V_{D1} = 60\eta$ mV. Thus, for $\eta = 2$, the change in $V_D$ needed to produce a 10:1 change in diode current is 120 mV; for $\eta = 1$, 60 mV is required.

small changes in $V_{AA}$, the portion of the diode characteristic between adjacent $Q$ points is approximately linear. However, for large changes in $V_{AA}$, as from $V_{AA1}$ to $V_{AA4}$, the portion of the diode curve is not linear. It is evident in Fig. 2-10*b* that an increase in $R$ causes $I_{DQ}$ to decrease.

## 2-7 LARGE-SIGNAL DIODE MODELS

It is convenient to represent the diode by a combination of ideal, linear circuit elements called an *equivalent circuit* or *circuit model*. As the diode is used with other circuit elements or devices, the model allows us to evaluate the currents and voltages in the network using standard circuit analysis methods.

The ideal diode (Fig. 2-7) is a *binary* device in the sense that it exists in only one of two possible states; that is, the diode is either ON or OFF at a given time. Consider a real diode such as that whose characteristic is shown in Fig. 2-11. If the applied voltage across this diode exceeds the cut-in voltage $V_\gamma$, with the anode $A$ (the $p$ side) more positive than the cathode $K$ (the $n$ side), the diode is forward-biased and is in the ON state. The OFF state exists when the applied voltage is less than $V_\gamma$ and, in effect, reverse-biases the diode.

As shown in Fig. 2-11*a*, the two line segments approximate the forward characteristic of the diode. This piecewise representation is modeled by a voltage source $V_\gamma$ in series with a resistance $R_f$ (usually 5 to 50 Ω for silicon diodes) as depicted in Fig. 2-11*b*. This piecewise linear characteristic has value because for $v_D < V_\gamma$, the forward current is sufficiently small that it can be neglected. Furthermore, the diode voltage drop is generally small in comparison with the applied voltages in the circuit so that the difference between the straight line and the actual characteristic introduces negligible error. In effect, the ON state can be regarded as an ideal diode in series with a battery $V_\gamma$ and a resistor $R_f$. It is important to note, however, that the only accessible terminals for measurement are $A$ and $K$.

For the OFF state, the diode characteristic is approximated by the straight line passing through the origin depicted in Fig. 2-12*a*, the slope of which is $1/R_r$. This representation gives rise to the equivalent circuit in Fig. 2-12*b*. As

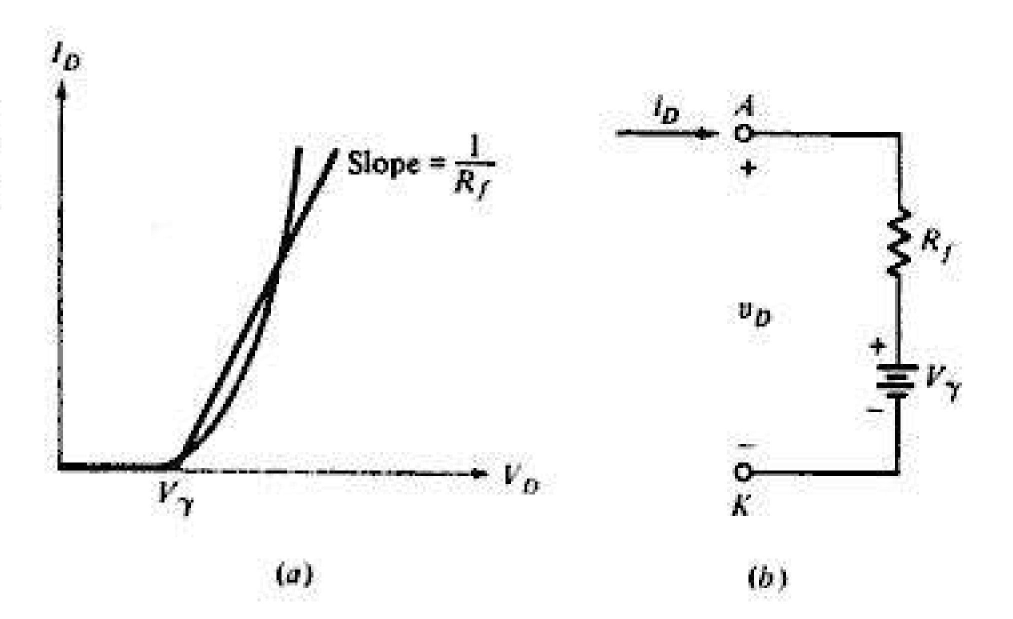
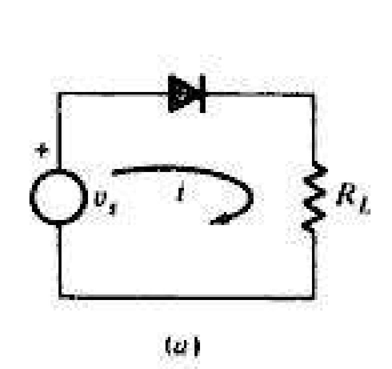
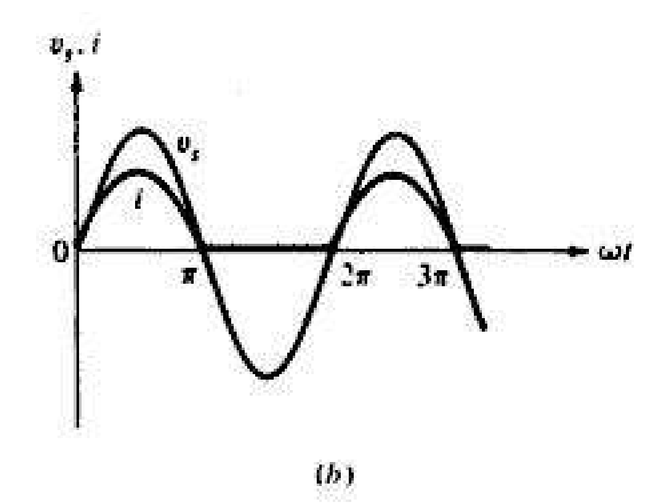



**FIGURE 2-11**
(*a*) Piecewise linear diode forward characteristic; (*b*) diode model for forward bias.

**FIGURE 2-15**
(*a*) Half-wave rectifier and (*b*) current and voltage waveforms.




## 2-8 ELEMENTARY DIODE APPLICATIONS

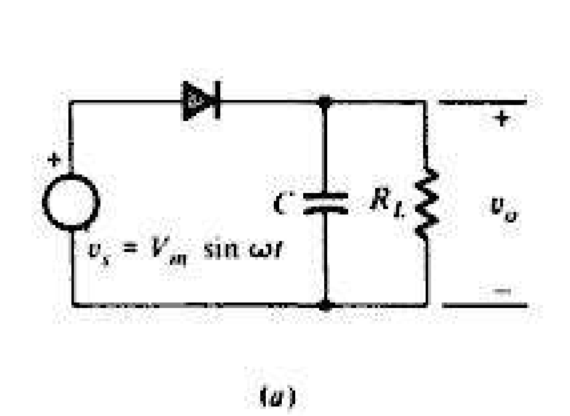
Several classes of circuits exploit the ON-OFF behavior of diodes to markedly alter electrical waveforms. We now introduce the basis of such circuits; specific applications are treated in subsequent chapters.

### Rectifiers

Consider the circuit in Fig. 2-15*a* in which the series combination of an ideal diode and load resistance is excited by a sinusoidal voltage. During the first half-cycle of the input the diode is ON and a current $v_s/R_L$ exists. The diode is OFF during the negative half-cycle of $v_s$ so that the current is zero (see Fig. 2-15). As current exists for only one-half of the cycle, the circuit in Fig. 2-15*a* is called a *half-wave rectifier*. It is significant that the average value (the dc component) of current over one period is not zero whereas the average voltage over one cycle is zero. This factor is the basis of the rectifier circuits used to convert the generally available alternating current to the direct current required by most electronic systems.

The circuit in Fig. 2-16*a* uses the capacitor $C$ as a simple filter to convert the waveform in Fig. 2-15*b* to the nearly constant (dc) level displayed in Fig. 2-16*b*. The following is a qualitative description of the effect of the capacitor on the circuit response. In the steady state at $t = t_1$, the voltage across the capacitor is $V_1$ and at that time the input voltage equals $V_1$, turning on the diode. Subsequent to $t_1$, the capacitor voltage, that is, the output voltage $v_o$,

**FIGURE 2-16**
(*a*) Rectifier with capacitor filter; (*b*) output voltage of circuit in part *a*.



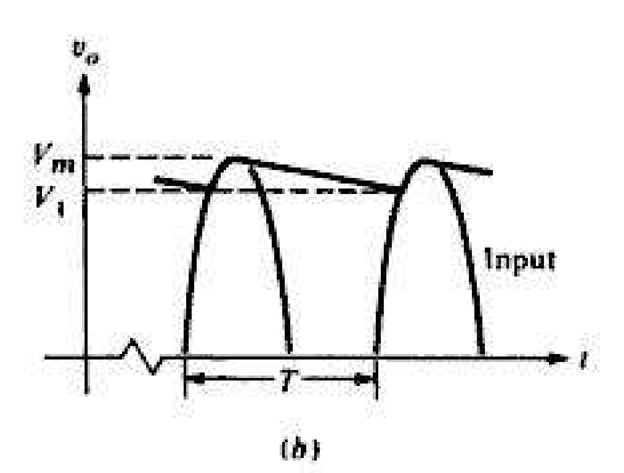
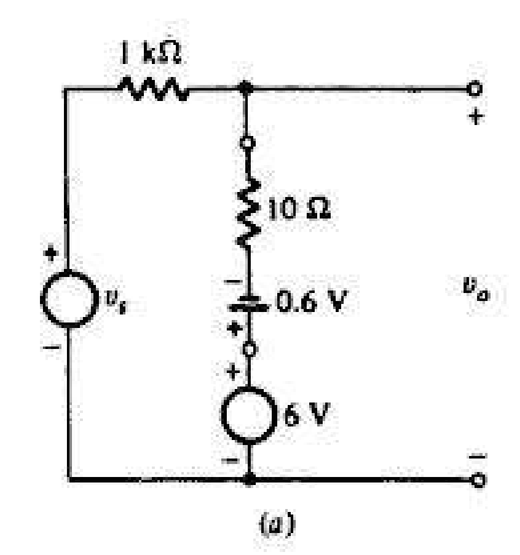
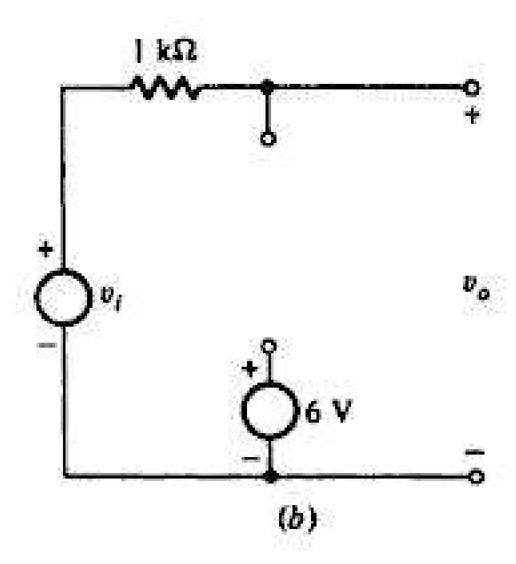
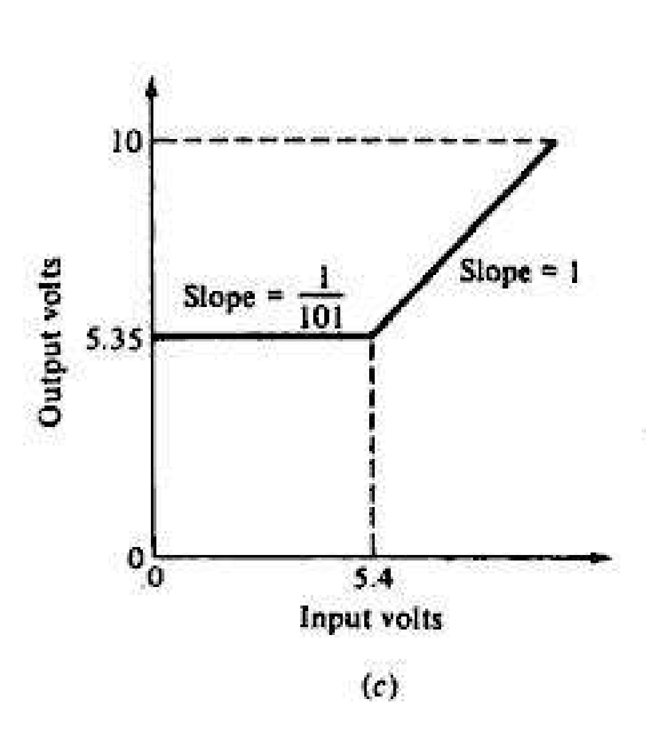
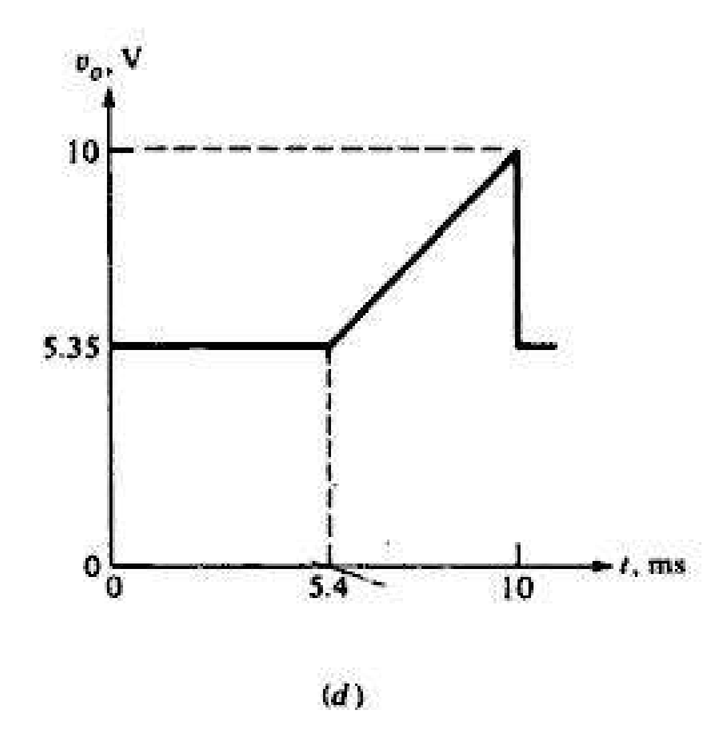



**FIGURE 2-22**
(*a*) Forward-bias and (*b*) reverse-bias models for Example 2-3; (*c*) transfer characteristic of circuit in Fig. 2-21*a*; (*d*) output waveform for the input in Fig. 2-21*b*.




The breakpoint is derived from the current relation and is the value of $v_s$ at which $i$ is zero. Thus

$$v_s = V_R - V_\gamma = 6 - 0.6 = 5.4 \text{ V}$$

The diode is ON for $v_s < 5.4$ V and OFF for $v_s > 5.4$ V. Substitution for $i$ in the equation for $v_o$ yields

$$v_o = \frac{R_f}{R_f + R} v_s + \frac{R(V_R - V_\gamma)}{R_f + R}$$

$$= \frac{10v_s}{10 + 1000} + \frac{1000(6 - 0.6)}{10 + 1000} = \frac{v_s}{101} + 5.35 \qquad \text{V}$$

The transfer characteristic is depicted in Fig. 2-22*c*; the output waveform, derived from the input waveform and transfer characteristic, is illustrated in Fig. 2-22*d*.

Three features of the output waveform can be discerned. First, the circuit in Fig. 2-21*a* is that shown in Fig. 2-18*a* with the diode connection reversed. As a result, the circuit in this example transmits input voltages above a given

Using the small-signal model in Fig. 2-26*b*, we obtain the alternating component of the output voltage from the voltage-divider relation as

$$v_{o,\mathrm{ac}} = \frac{2000}{2000 + 12} \times 0.2 \sin \omega t = 0.199 \sin \omega t \qquad \mathrm{V}$$

(*b*) The total load voltage $v_o(t)$ is the sum of the quiescent and time-varying components. The quiescent voltage across $R_L$ is

$$V_{OQ} = I_{DQ}R_L = 4.18 \times 10^{-3} \times 2 \times 10^3 = 8.36 \mathrm{~V}$$

Thus

$$v_o(t) = 8.36 + 0.199 \sin \omega t \qquad \mathrm{V}$$

The waveforms in Fig. 2-27 are what we would observe if the output were displayed on an oscilloscope. In Fig. 2-27*a*, with the selector knob set on dc, the oscilloscope displays $v_o$, the total instantaneous output voltage. We observe that the alternating component is almost indistinguishable. However, in Fig. 2-27*b* with the selector knob set to ac and the sensitivity increased, the alternating component is conveniently displayed for measurement. In effect, the small-signal model performs an analogous function to switching the selector knob from dc to ac. By removing the quiescent (dc) level, we can focus on the effect that the time-varying input has on the time-varying response.

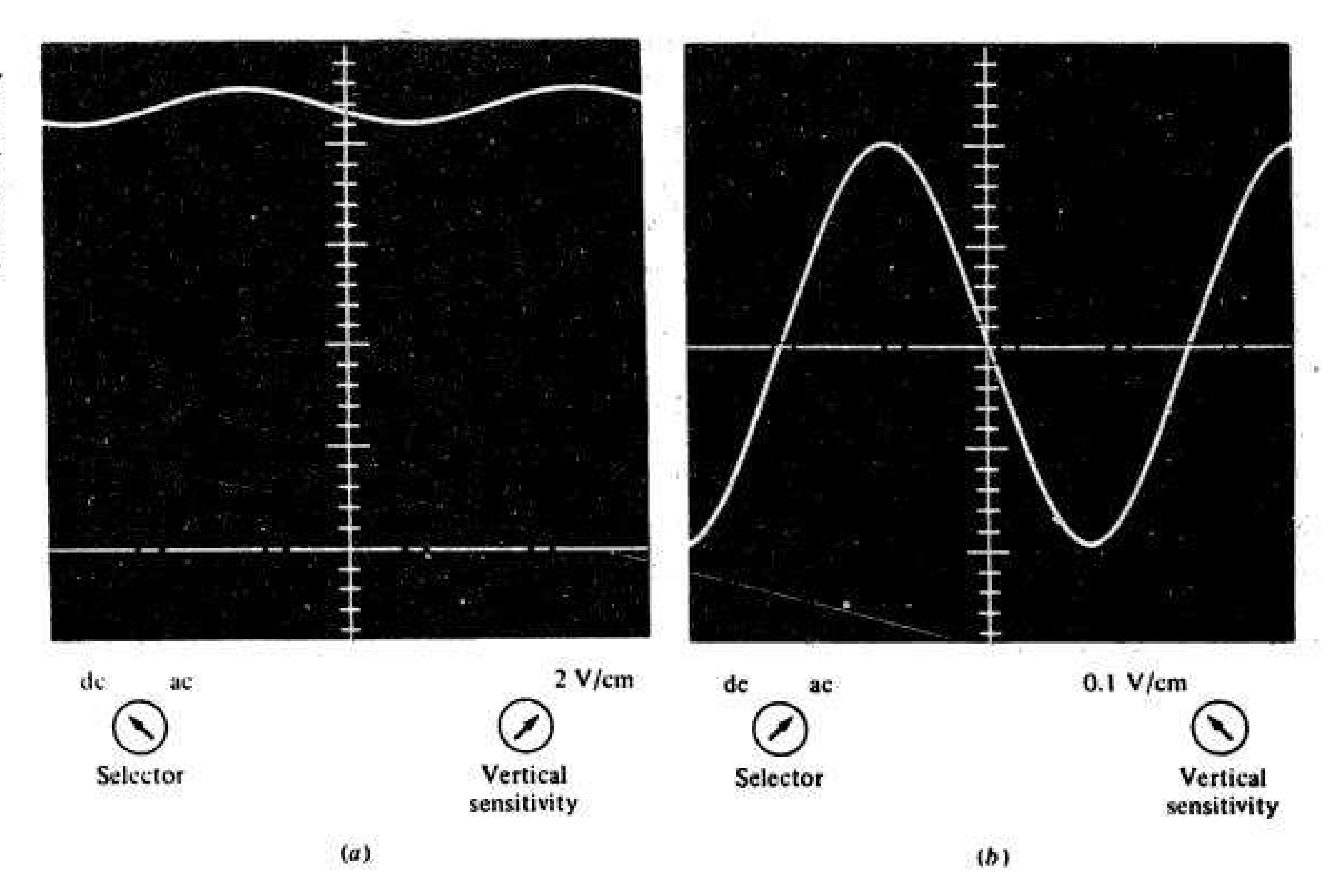



**FIGURE 2-27**
Oscilloscope displays of output voltage in Example 2-4: (*a*) selector knob set to dc and (*b*) selector knob set to ac. The dashed line indicates 0 V.

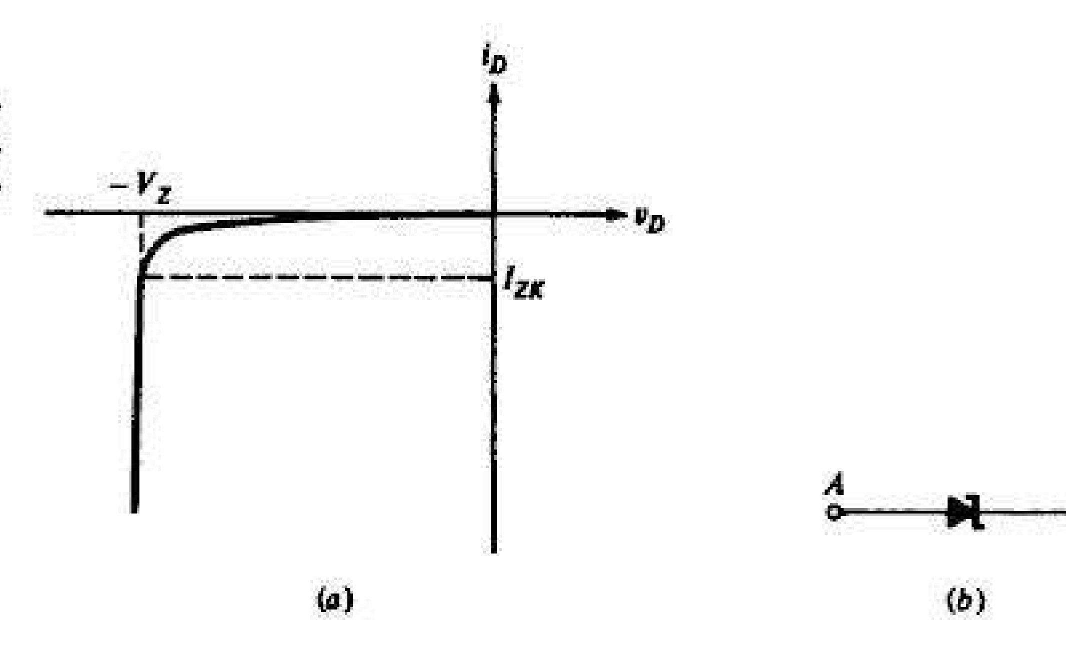


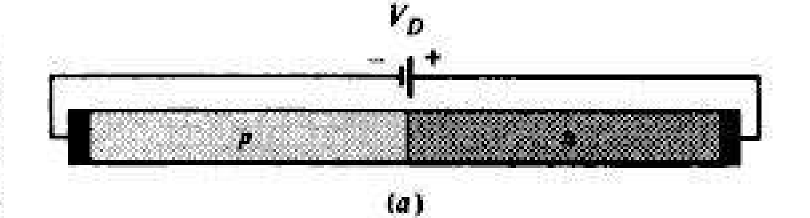

**FIGURE 2-30**
(*a*) Reverse-bias characteristic showing breakdown region; (*b*) Zener diode symbol.

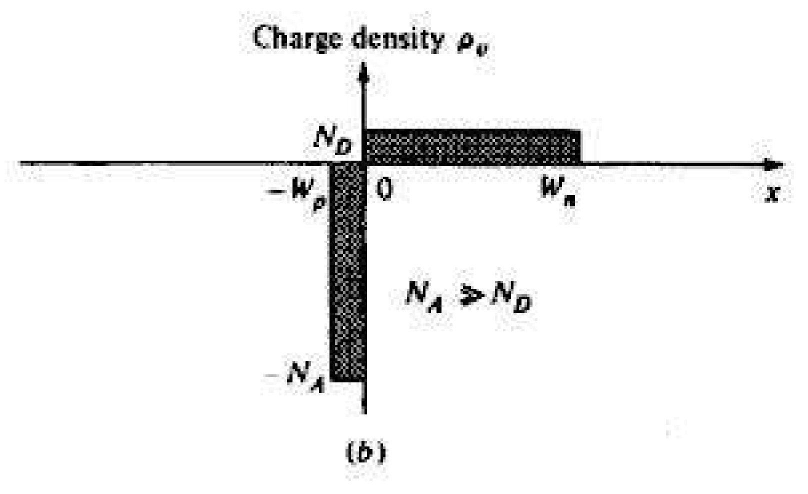
sipation capabilities to operate in the breakdown region are commonly called *Zener diodes*, whose circuit symbol is indicated in Fig. 2-30*b*. (The term "Zener diode" is used independently of the breakdown mechanism.) These devices are employed as voltage regulators and in other applications in which a constant reference voltage is required.

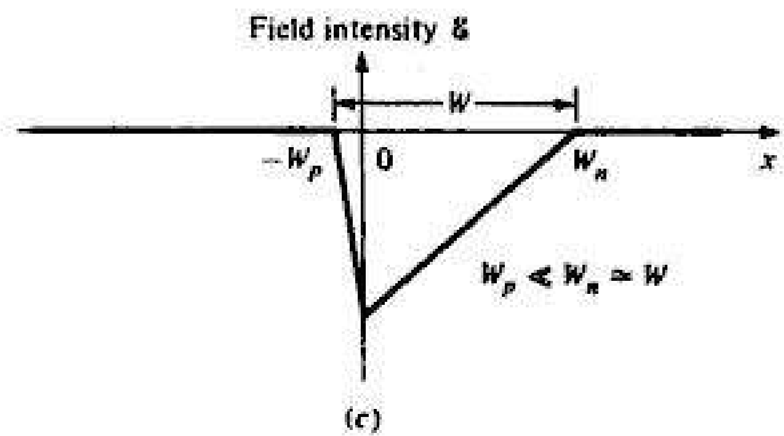
**Avalanche Multiplication** Consider the following situation in a reverse-biased diode. A thermally generated carrier (part of the reverse saturation current) falls down the junction barrier and acquires energy from the applied potential. This carrier collides with a crystal ion and imparts sufficient energy to disrupt a covalent bond. In addition to the original carrier, a new electron-hole pair has now been generated. These carriers may also pick up sufficient energy from the applied field, collide with another crystal ion, and create still other electron-hole pairs. Thus each new carrier may, in turn, produce additional carriers through collision and the action of disrupting bonds. This cumulative process is referred to as *avalanche multiplication*. It results in large reverse currents, and the diode is said to be in the region of *avalanche breakdown*.

**Zener Breakdown** Even if the initially available carriers do not acquire sufficient energy to disrupt bonds, it is possible to initiate breakdown through a direct rupture of the bonds. Because of the existence of the electric field at the junction, a sufficiently strong force may be exerted on a bound electron by the field to tear it out of its covalent bond. The new hole-electron pair which is created increases the reverse current. Note that this process, called *Zener breakdown*, does not involve collisions of carriers with the crystal ions.

The field intensity $\mathscr{E}$ increases as the impurity concentration increases, for a fixed applied voltage. It is found that Zener breakdown occurs at a field of approximately $2 \times 10^7$ V/m. This value is reached at or below about 6 V for heavily doped junctions. For lightly doped diodes, the breakdown voltage is higher and avalanche multiplication is the predominant effect. Silicon diodes operated in avalanche breakdown are available with maintaining voltages from several volts to several hundred volts and with power ratings up to 50 W.



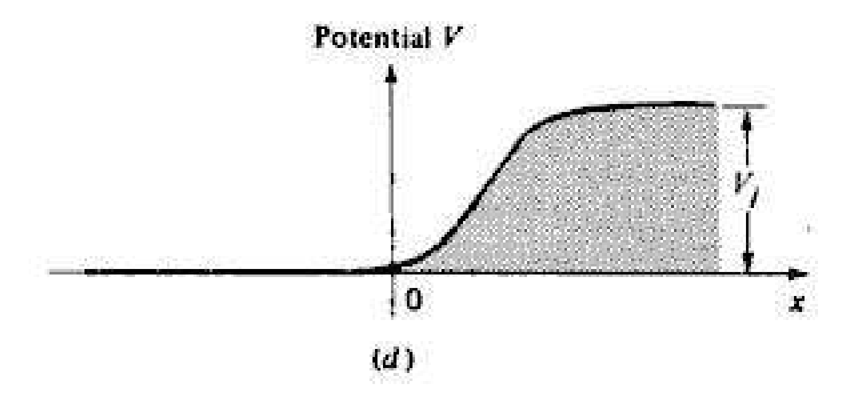




**FIGURE 2-34**
(a) A reverse-biased step-graded pn junction; (b) the charge density, (c) the field intensity; (d) the potential variation with distance x from the junction.

If $N_A \gg N_D$, then $W_p \ll W_n \simeq W$. The relationship between potential and charge density is given by Eq. (2-1):

$$\frac{d^2V}{dx^2} = \frac{-qN_D}{\epsilon} \qquad \text{V/m}^2 \tag{2-24}$$

The electric lines of flux start on the positive donor ions and terminate on the negative acceptor ions. Hence there are no flux lines to the right of the boundary $x = W_n$ in Fig. 2-34, and $\mathscr{E} = -dV/dx = 0$ at $x = W_n \simeq W$. Integration of Eq. (2-24) subject to this boundary condition yields

$$\frac{dV}{dx} = \frac{-qN_D}{\epsilon}(x - W) = -\mathscr{E}(x) \qquad \text{V/m} \tag{2-25}$$

where Eq. (2-32) is used for $p_n(x)$. The hole current $I$ is given by $I_p(x)$ in Eq. (2-36), with $x = 0$, or

$$I = \frac{AqD_p p'(0)}{L_p} \quad \text{A} \tag{2-37}$$

The excess minority charge $Q$ exists only on the $n$ side and is given by the shaded area in the $n$ region in Fig. 2-35$a$ multiplied by the diode cross section $A$ and the electronic charge $q$. Hence, from Eq. (2-32), we obtain

$$Q = \int_0^\infty Aqp'(0)\,\epsilon^{-x/L_p}\,dx = AqL_p p'(0) \quad \text{C} \tag{2-38}$$

Elimination of $p'(0)$ from Eqs. (2-37) and (2-38) yields

$$I = \frac{Q}{\tau} \quad \text{A} \tag{2-39}$$

where $\tau \equiv L_p^2/D_p \equiv \tau_p$ = mean lifetime for holes [Eq. (2-33)].

Equation (2-38) is an important relationship, referred to as the *charge-control description of a diode*. It states that the diode current (which consists of holes crossing the junction from the $p$ side to the $n$ side) is proportional to the stored charge $Q$ of excess minority carriers. The factor of proportionality is the reciprocal of the decay time constant (the mean lifetime $\tau$) of the minority carriers. Thus, in the steady state, *the current $I$ supplies minority carriers at the rate at which these carriers are disappearing because of the process of recombination.*

The charge-control characterization of a diode describes the device in terms of the current $I$ and the stored charge $Q$, whereas the equivalent-circuit characterization uses the current $I$ and the junction voltage $V$. One immediately apparent advantage of this charge-control description is that the exponential relationship between $I$ and $V$ is replaced by the linear dependence of $I$ on $Q$. The charge $Q$ also makes a simple parameter, the sign of which determines whether the diode is forward- or reverse-biased. The diode is forward-biased if $Q$ is positive and reverse-biased if $Q$ is negative.

**Diffusion Capacitance** In Sec. 2-9 we introduced the diffusion capacitance $C_D$ as a means for modeling minority-carrier storage in the vicinity of a forward-biased diode. We can now derive this element based on the charge-control description just concluded. From Eqs. (2-39) and (2-17)

$$C_D \equiv \frac{dQ}{dV} = \tau\frac{dI}{dV} = \tau g_d = \frac{\tau}{r_d} \quad \text{F} \tag{2-40}$$

where the diode incremental conductance is $g_d \equiv dI/dV$. Substitution of the expression for the diode incremental resistance $r_d = 1/g_d$ given in Eq. (2-18) into Eq. (2-40) yields

$$C_D = \frac{\tau I_D}{\eta V_T} \quad \text{F} \tag{2-41}$$

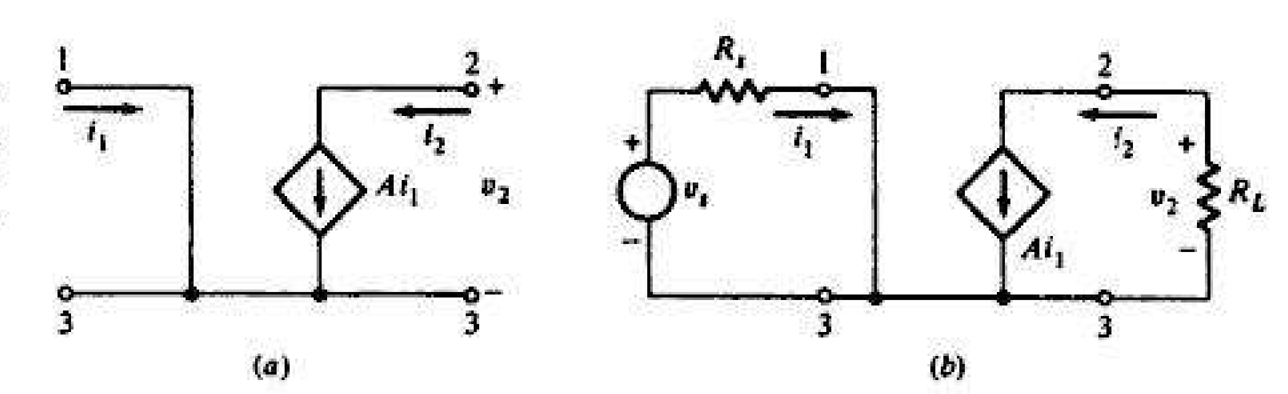



**FIGURE 3-1**
(*a*) Ideal current-controlled current source with (*b*) voltage excitation and load resistance.

a current source of strength $Ai_1$ acts as the output terminal pair (2–3).[1] Parameter $A$ relates the strength of the source to the control current and is often referred to as the *current gain*. Physically, $A$ is related to the processes which occur within the device used to obtain the controlled source.

It is evident in Fig. 3-1*a* that the effect of $i_1$ is transmitted to the output by means of the source, whereas signals applied at the output terminals do not affect the control current. This *unilateral* behavior allows the portions of a circuit where the control signal is applied to be isolated from circuit elements connected to the output.

In the circuit shown in Fig. 3-1*b*, a signal source $v_s$ in series with a resistance $R_s$ is connected to the input and a load resistance $R_L$ is placed across the output of the controlled source. The control current is $i_1 = v_s/R_s$, and the output voltage $v_2$ is

$$v_2 = -Ai_1R_L = -\frac{AR_L}{R_s}v_s \tag{3-1}$$

From Eq. (3-1) we see that when $|AR_L/R_s| > 1$, then $|v_2| > |v_s|$ and voltage amplification is achieved. In addition, if $A > 1$ as is usually the case, current gain is realized as the output current is greater than the input current. It is also true that the power dissipated by $R_L$ is greater than the power supplied by $v_s$. Thus the controlled source is capable of providing power gain. One consequence of amplification is that the power required for control is less than the amount of power controlled. This, coupled with the unilateral property, enables the controlled source to be operated as a controlled switch.[2]

The output volt-ampere characteristics are a convenient method for displaying the dependence of source strength on the control variable. These char-

[1] A controlled source is designated by the diamond symbol as is depicted in Fig. 3-1. For a current source, the arrow in the diamond indicates the direction of the current, whereas the + and − signs placed at the top and bottom of the diamond indicate the polarity of the voltage source. In either case, the strength of the source is shown algebraically adjacent to the diamond.

[2] We are aware that it is desirable to control substantial quantities of energy with a small expenditure of energy. Consider the mechanical energy necessary to turn the switch which activates a 120-V 10-A window-mounted air conditioner. Roughly, this requires that a 0.05-kilogram (kg) switch be moved 1 cm in 0.25 s and is about $5 \times 10^{-3}$ J at a rate of 0.02 W. Yet the energy and power controlled by this effort are equivalent to lifting the unit from the floor to the window sill in 0.25 s.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

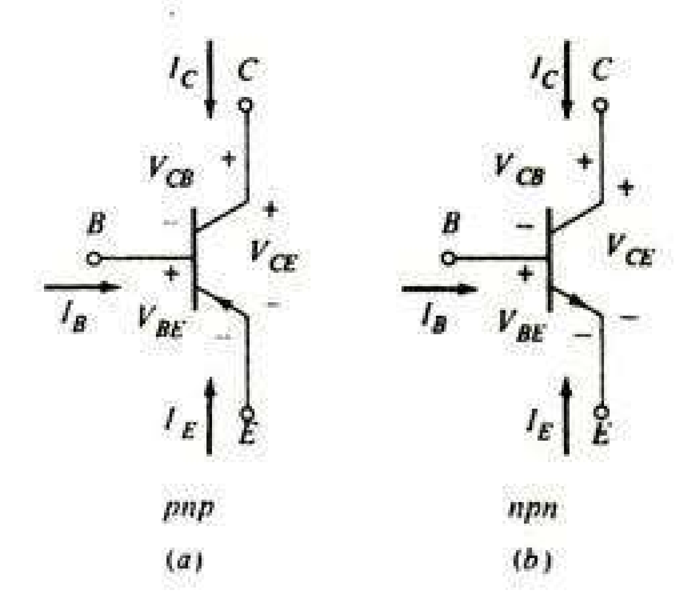
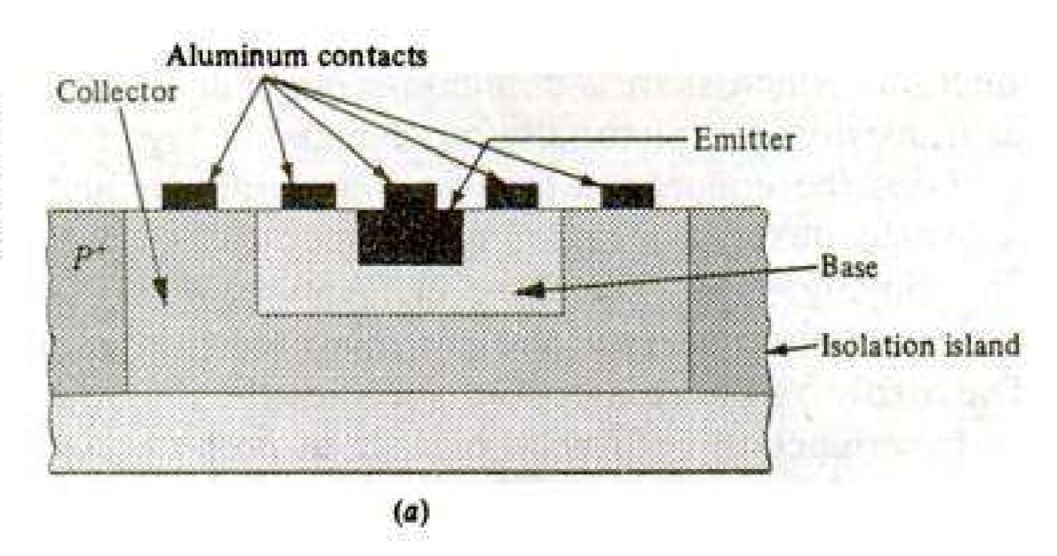



**FIGURE 3-6**
Circuit symbols for (*a*) *pnp* and (*b*) *npn* transistors.



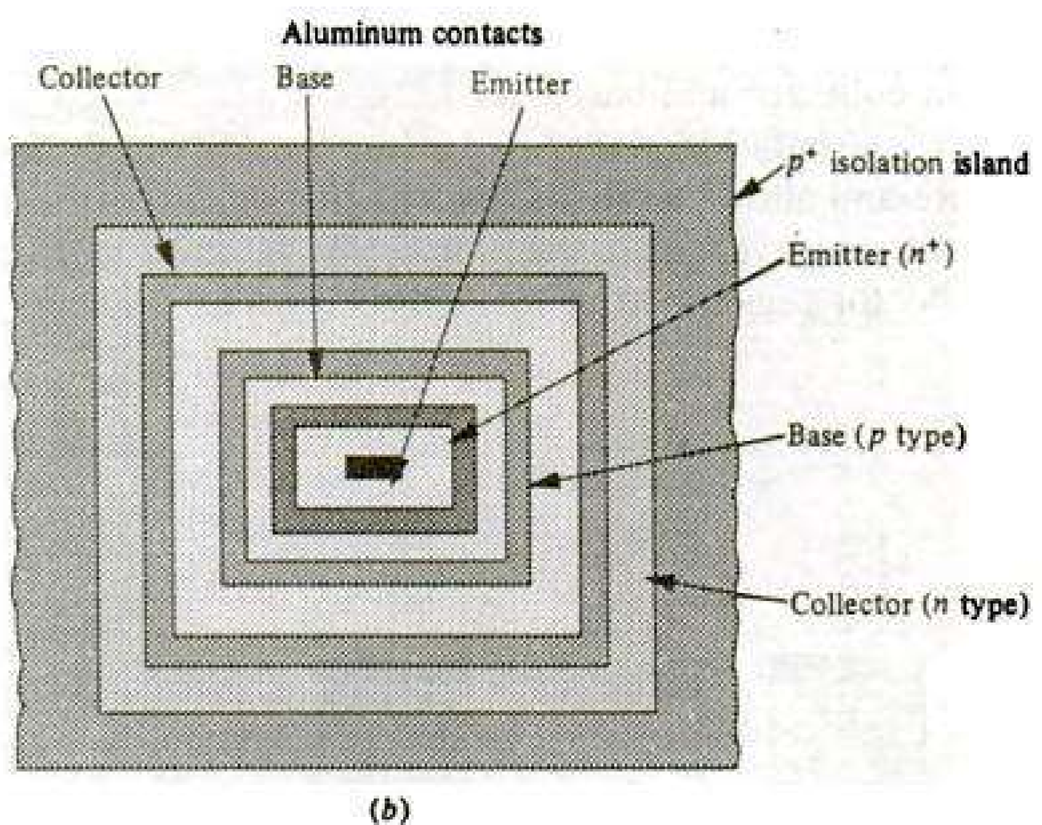




**FIGURE 3-7**
Structure of an IC *npn* transistor including *p* substrate and isolation island and aluminum contacts.

of the transistor with shading used to represent the several impurity concentrations. In the top view in Fig. 3-7*b* it is shown how aluminum contacts are made to the collector, base, and emitter regions. Note that because of the selective doping of the silicon block, a *p* region is sandwiched between two *n*

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

given in Eq. (2-3).[1] Thus $I_E$ and $I_C$ can be expressed in terms of the two diode currents as

$$I_E = I_{ED} - \alpha_R I_{CD} = I_{ES}(\epsilon^{V_{EB}/V_T} - 1) - \alpha_R I_{CS}(\epsilon^{V_{CB}/V_T} - 1) \tag{3-2}$$

$$I_C = -\alpha_F I_{ED} + I_{CD} = -\alpha_F I_{ES}(\epsilon^{V_{EB}/V_T} - 1) + I_{CS}(\epsilon^{V_{CB}/V_T} - 1) \tag{3-3}$$

The relationships expressed in Eqs. (3-2) and (3-3) are known as the *Ebers-Moll equations*.

The quantities $I_{ES}$ and $I_{CS}$ in Eqs. (3-2) and (3-3) are the reverse saturation currents of the emitter-base and collector-base junctions, respectively. The parameters $\alpha_F$ and $\alpha_R$ are each less than unity, as not all the current from one diode is coupled to the other junction. The subscripts refer to forward ($F$) transmission from emitter to collector and reverse ($R$) transmission from collector to emitter. The four quantities $I_{ES}$, $I_{CS}$, $\alpha_F$, and $\alpha_R$ are functions of doping densities and transistor geometry. These quantities are not independent but are found from theory to be related by

$$\alpha_F I_{ES} = \alpha_R I_{CS} \tag{3-4}$$

This equation is often called the *reciprocity condition* for the BJT.

The base current is obtained by realizing that the sum of the terminal currents must be zero (KCL). Thus

$$I_B = -(I_E + I_C) \tag{3-5}$$

It is instructive to indicate the typical values of the quantities in the Ebers-Moll equations. For IC transistors (Fig. 3-7) of the nominally small dimensions used

$$0.98 \le \alpha_F \le 0.998 \qquad \text{and} \qquad 0.40 \le \alpha_R \le 0.8$$

and $I_{ES}$ and $I_{CS}$ are in the order of $10^{-15}$ A. Both $I_{ES}$ and $I_{CS}$ depend on their respective junction areas. Consequently, for specified donor and acceptor doping levels, transistor currents can be scaled by changing device dimensions. This is typically done in IC design to realize transistors with differing current-handling capability. By this method $I_{ES}$ and $I_{CS}$ can be increased to about $10^{-13}$ and $10^{-12}$ A, respectively. The value of $\alpha_F$ remains essentially unchanged, and [from Eq. (3-4)] $\alpha_R$ can be reduced below 0.1. Scaling device dimensions are utilized in discrete transistors to achieve higher current and power levels than are possible on a chip.

The Ebers-Moll equations for an *npn* device are obtained from Eqs. (3-2) and (3-3) once we recognize that the forward current in each diode is from $p$ to $n$, and that forward bias requires a positive voltage from $p$ to $n$. Consequently, the directions of all current components and junction voltages for an *npn* transistor are reversed from those for a *pnp* device as depicted in Fig.

[1] Most IC transistors operate at currents which are typically at least nine orders of magnitude greater than saturation currents. Consequently, $\eta = 1$ as described in Sec. 2-3.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

where the reciprocity condition, $\alpha_R I_{CS} = \alpha_F I_{ES}$ from Eq. (3-4) has been used. Inverting and taking the logarithm of both sides, we have

$$\frac{V_{EB}}{V_T} = \ln \frac{1}{1 - \alpha_F} = \ln (\beta_F + 1) \tag{2}$$

In (2) we observe that $V_{EB}$ is positive, thus reverse-biasing the emitter junction. Consequently, the transistor is cut off.

(*b*) With $I_E = 0$, KCL requires $I_C = -I_B$. The collector current is obtained from Eq. (3-7), into which (1) is substituted:

$$I_C = -I_B = -\alpha_F \alpha_R I_{CS} + I_{CS} = (1 - \alpha_F \alpha_R) I_{CS} \tag{3}$$

(*c*) Substitution of values in (2) gives

$$\frac{V_{EB}}{25 \times 10^{-3}} = \ln \frac{1}{1 - 0.99} \qquad \text{and} \qquad V_{EB} = 115 \text{ mV}$$

The value of $\alpha_R$, from the reciprocity condition, is

$$\alpha_R = \alpha_F \frac{I_{ES}}{I_{CS}} = 0.99 \frac{10^{-15}}{2 \times 10^{-15}} = 0.495$$

Then, using (3), we obtain

$$I_C = -I_B = (1 - 0.99 \times 0.495) \times 2 \times 10^{-15} = 1.02 \times 10^{-15} \text{ A}$$

With $I_E = 0$, the result indicates that the transistor, between the base and collector terminals, behaves as a diode and the current determined is the effective reverse saturation collector current for an open-circuited emitter. Although the value of $I_C$ obtained is small, it increases markedly with temperature.

The current given by (3) in Example 3-1 is often referred to as the *reverse collector current*. As we see in the next section, this is an important quantity in BJTs and is usually designated by $I_{CO}$. Using a similar analysis, with the collector open-circuited and the emitter-base diode reverse-biased, the reverse emitter current $I_{EO}$ is obtained. The two results are stated in Eq. (3-15):

$$I_{CO} = (1 - \alpha_F \alpha_R) I_{CS} \qquad I_{EO} = (1 - \alpha_F \alpha_R) I_{ES} \tag{3-15}$$

**Minority-Carrier Concentrations** The excess minority-carrier concentration in the base region, due primarily to forward injection, is depicted in Fig. 3-12. The width $W$ of the base region is defined as the distance between the base side of the emitter-base and collector-base depletion regions. Ideally, the excess minority-carrier density decreases linearly across the base region. In reality, the distribution is given by the dashed line, which takes recombination into account. The concentration is zero at the collector-base boundary as minority carriers reaching this point are swept into the collector.

At a constant value of $V_{EB}$, the Early effect predicts that $I_E$ increases as we increase $|V_{CB}|$. This conclusion accounts for the shift in the input characteristics shown in Fig. 3-14. In Sec. 3-5 we observe other manifestations of base-width modulation.

**3-5 THE COMMON-EMITTER (*CE*) CONFIGURATION** Most BJT circuits employ the common-emitter configuration shown in Fig. 3-15. This is due mainly to the fact that it is desirable to use the small base current as the control quantity rather than the emitter current. In the *CE* configuration, $I_B$, the input current, and $V_{CE}$, the output voltage, are the independent variables, whereas the input voltage $V_{BE}$ and the output current $I_C$ are the dependent variables.

We feel that the physical operation of the BJT is somewhat easier to understand if reference is made to *pnp* devices. Hence the preceding discussions of the *CB* configuration and the Ebers-Moll equations focused on the *pnp* transistor. However, *npn* devices are most prevalent in both ICs and discrete-component circuits which employ BJTs. Our discussion of the *CE* configuration thus concentrates on *npn* transistors, and, as previously stated, we utilize the 2N2222A, a discrete transistor which is a widely used industry standard.

**The Output Characteristics**[1] The common-emitter output characteristics is the family of curves shown in Fig. 3-16 in which $I_C$ versus $V_{CE}$ is plotted for various values of $I_B$. A load line corresponding to $R_C = 500\ \Omega$ and a supply voltage $V_{CC} = 10$ V has been superimposed on these characteristics. Construction of the load line is based on KVL for the output loop and is identical to the method explained in Sec. 2-4. The output characteristics display three regions of operation, as did the common-base characteristics. The active region is discussed here; cutoff and saturation are considered in the next section.

In the active region, for an *npn* transistor, Eq. (3-17) must be modified to $I_C = -\alpha_F I_E + I_{CO}$. Combination of this equation with Eq. (3-5) yields

[1]Transistor output and input characteristics are no longer supplied by the transistor manufacturer since they are seldom used in either digital or analog design. However, these characteristics are necessary for an understanding of the transistor. The device characteristics shown in this chapter were obtained experimentally.

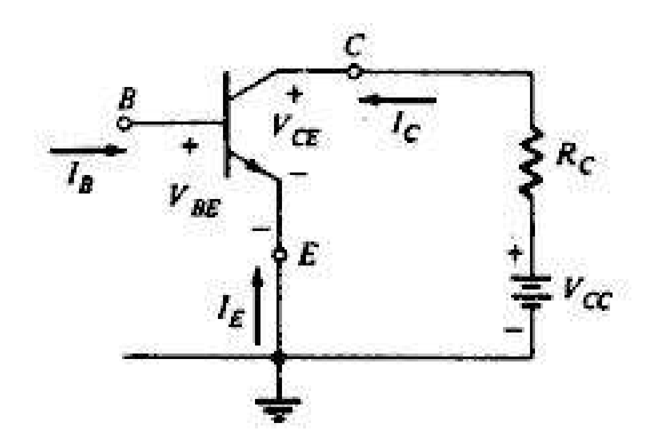



**FIGURE 3-15**
A common-emitter circuit employing an *npn* transistor.

biased, the input characteristic is essentially that of a forward-biased diode. If $V_{BE}$ becomes zero, then $I_B$ will be zero, since under these conditions both emitter and collector junctions will be short-circuited. In general, increase in $|V_{CE}|$ with constant $V_{BE}$ causes a decrease in base width $W$ due to the Early effect and results in a decreasing recombination base current. These considerations account for the shape of input characteristics shown in Fig. 3-20.

**The Reverse-Active Mode** The input and output characteristics of the inverted transistor have the same general shape as in Figs. 3-20 and 3-16. The reverse-active input characteristics display the behavior of the forward-biased collector-base junction. Recall that in the reverse-active mode, $\alpha_R$ and $\beta_R$ have lower values than $\alpha_F$ and $\beta_F$, respectively. Consequently, for a given value of $I_B$, lower values of $I_E$ exist in the reverse-active mode than in the forward-active region.

## 3-6 CUTOFF AND SATURATION MODES

We indicated qualitatively in Sec. 3-2 that operation of a BJT in cutoff and saturation approximates the state of an open and closed switch, respectively. In this section we investigate the behavior of the transistor in these modes in more quantitative terms.

### Cutoff

Both junctions are reverse-biased in cutoff. In the *CB* configuration we showed that cutoff occurs when the input current $I_E = 0$ and, consequently, $I_C = -I_B = I_{CO}$. For the *CE* configuration, we now examine operation when the input current $I_B = 0$. It is important to note that, theoretically, both junctions are not reverse-biased if the base is open-circuited (Prob. 3-5). From Eq. (3-5), if $I_B = 0$, $I_E = -I_C$, and use of Eq. (3-18) gives

$$I_C = -I_E = \frac{I_{CO}}{1 - \alpha_F} \equiv I_{CEO} \tag{3-22}$$

The rationale for the subscripts of the current $I_{CEO}$ in Eq. (3-22) is that this is the current from *C* to *E* when *B* (the missing terminal) is *O*pen. At $I_C \approx I_{CO}$, $\alpha_F$ is very nearly zero because of recombination in the emitter-base depletion region. Hence, from Eq. (3-22), we find that $I_C = -I_E = I_{CEO} = I_{CO}$, and, for practical purposes, the BJT is very nearly cut off. In Example 3-1,[1] the relationship given by Eq. (2) indicates that for $\alpha_F$ approaching zero, $V_{EB} \approx 0$. Thus cutoff ($I_E = 0$) *for a silicon transistor requires a reverse voltage* $V_{BE}$ *that is virtually zero and* $-I_B = I_C = I_{CO}$.

The collector current in a physical transistor (a real, nonidealized, commercial device) when the emitter current is zero is designated by the symbol $I_{CBO}$. Two factors cooperate to make $I_{CBO}$ larger than $I_{CO}$: (1) there exists a leakage current which flows, not through the junction, but around it and across the surfaces—this current is proportional to the voltage across the junction;

[1] In this example, we assumed that $\alpha_F$ was constant at all current levels.

**TABLE 3-3 Typical Junction Voltages at 25°C**

| *Quantity* | $V_{CE}$ *at edge of saturation* | $V_{CE(sat)}$ | $V_{BE}$ *Cut-in* | $V_{BE}$ *Active* | $V_{BE}$ *Saturation* | $V_{BE}$ *Cutoff* |
|---|---|---|---|---|---|---|
| Value (in V) | 0.3 | 0.2 | 0.5 | 0.7 | 0.8 | 0 |

It is reasonable to expect that the temperature variation of the voltage across a forward-biased junction is the same as that for a diode, namely, −2.2 mV/°C. In saturation the transistor consists of two forward-biased diodes back to back in series opposing. Hence it is to be anticipated that the temperature-induced voltage change in one junction will be canceled by the change in the other junction. We do indeed find such to be the case for $V_{CE(sat)}$ whose temperature coefficient is about one-tenth that of $V_{BE(sat)}$.

The values of circuit currents and voltages obtained from "pencil-and-paper" calculations based on the data in Table 3-3 correlate well with experimental values. The reader should be aware, however, that these values are typical and not exact. A variety of reasons exist in the design, fabrication, and manufacture of circuits which require the designer to have more accurate results. In these situations, computer simulations, such as SPICE, are widely employed. Pencil-and-paper calculations are still used to indicate nominal values of circuit quantities.

## 3-7 DC MODELS

We can construct a dc model for each operating region of the BJT from the previous discussions of the Ebers-Moll equations. Our focus is on the *CE* configuration, but the models apply equally well to the *CB* circuit.

The model for the forward-active region is displayed in Fig. 3-23*a* and is based on Eq. (3-11). Because reverse saturation currents are generally negligibly small, their effect is usually omitted. The battery in the base-emitter circuit is $V_{BE}$ and, from Table 3-3, is usually 0.7 V. The controlled current source $\beta_F I_B$ relates $I_C$ and $I_B$ in the active region. The resistance $R_0$, indicated by the dashed symbol, is the consequence of the Early effect. Generally, $R_0$ is sufficiently

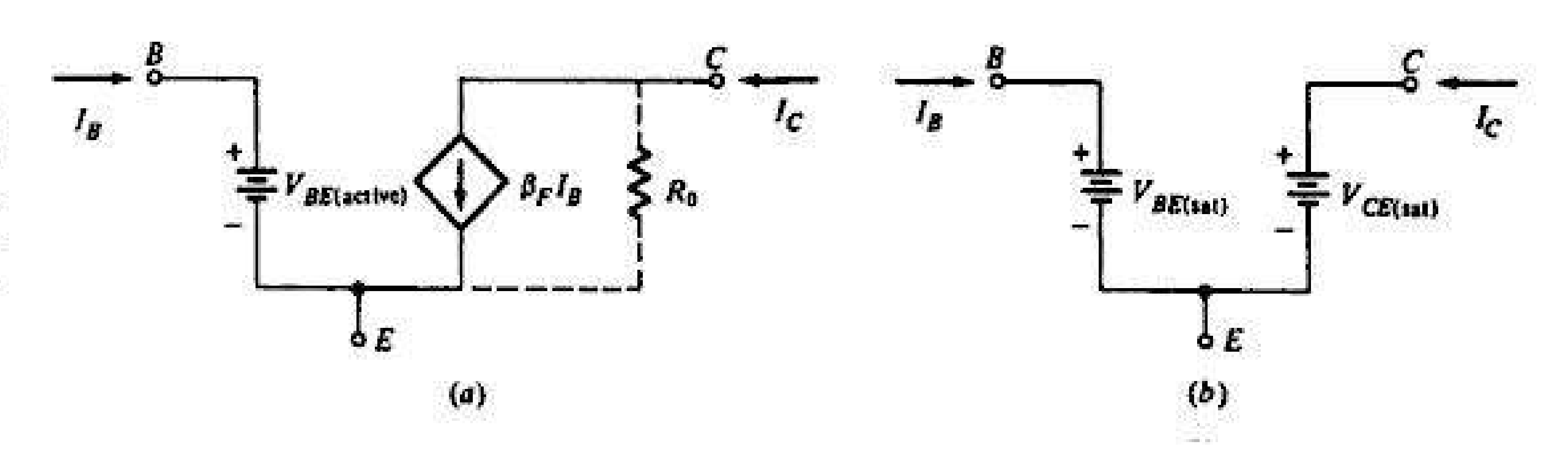



**FIGURE 3-23** Large-signal (dc) equivalent circuits for an *npn* transistor for (*a*) forward-active and (*b*) saturation-region operation.

Obviously, $V_{CE} > 0.3$ V and verifies our assumption that the operation is in the forward-active region. Hence

$$I_C = \beta_F I_B = 100 \times 0.0251 = 2.51 \text{ mA}$$

***Example 3-4*** (*a*) Determine $I_C$ and $V_{CE}$ for the circuit in Fig. 3-26*a*. The transistor has $\beta_F = 150$. (*b*) What is the minimum value of $R_C$ for which the transistor is just barely saturated?

***Solution*** (*a*) For convenience, Fig. 3-26*a* is redrawn as is shown in Fig. 3-26*b*. The base-bias network indicated in Fig. 3-26*b* can be replaced by its Thévenin equivalent and is so indicated in Fig. 3-26*c*, where

$$V_{BB} = \frac{R_2}{R_1 + R_2} V_{CC} = \frac{11 \times 12}{110 + 11} = 1.09 \text{ V}$$

$$R_B = R_1 \parallel R_2 = \frac{R_1 R_2}{R_1 + R_2} = \frac{110 \times 11}{110 + 11} = 10 \text{ k}\Omega$$

Note the similarity of this circuit with that in Fig. 3-24*a* (with the addition of $R_E$). We will again assume operation in the forward-active region for which the

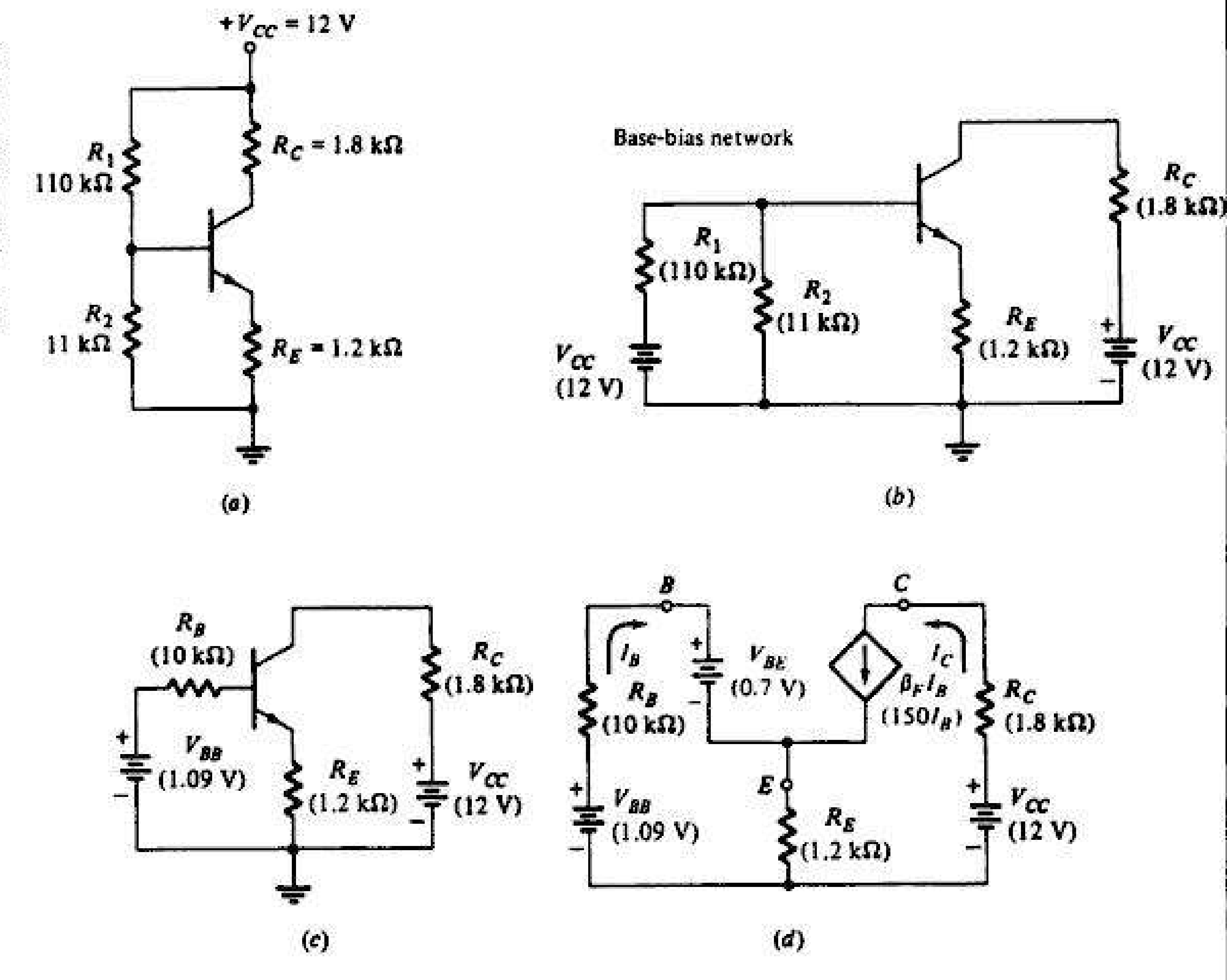
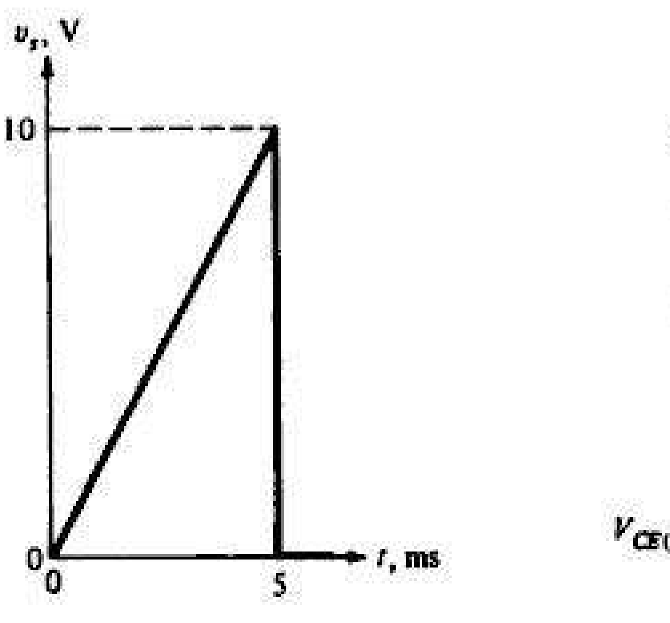



**FIGURE 3-26** (*a*) Circuit for Example 3-4; (*b*) circuit in part *a* showing base-bias network. The base-bias network is replaced by its Thévenin equivalent in part *c*. The transistor is represented in its forward-active model in part *d*.



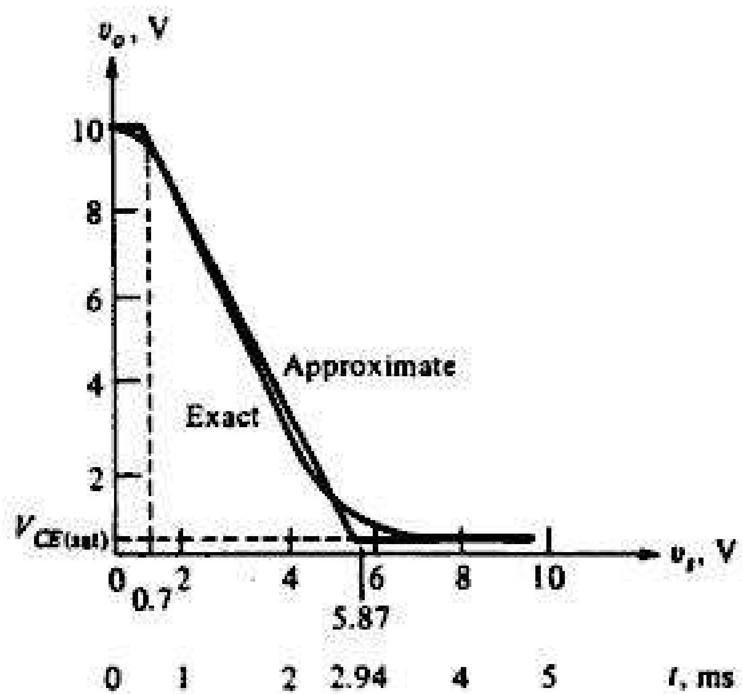




**FIGURE 3-29** (*a*) The input waveform and (*b*) transfer characteristic for Example 3-6.

voltage is exceeded. In Fig. 3-16 for $I_B \approx 0$, we see that $v_o = v_{CE} = 10$ V. We indicated in Sec. 3-5 that increasing $V_{CE}$ shifts the curves in Fig. 3-20*b* to the right. Thus we can assume, for $V_{CE} = 10$ V, that cut in occurs near $V_{BE} = 0.7$ V. This is expected as increasing $v_s$ causes a transition in BJT operation from cutoff to the forward-active region.

Once the transistor is in the active region, $V_{BE} = 0.7$ V and, from the KVL expression for the base-emitter loop,

$$i_B = \frac{v_s - V_{BE}}{R_B} = \frac{v_s - 0.7}{47} \text{ mA}$$

As $i_B$ increases linearly with $v_s$, we see that along the load line $v_{CE}$ decreases almost linearly until the transistor approaches saturation. The onset of saturation (shown in Fig. 3-16) occurs as $I_B$ approaches 120 $\mu$A. Let us interpolate this value of $I_B$ from Fig. 3-16 as 110 $\mu$A. The corresponding value of $v_s$ is

$$0.110 = \frac{v_s - 0.7}{47} \quad \text{and} \quad v_s = 5.87 \text{ V}$$

Further increase in $v_s$, and hence $i_B$, has no effect on the output.

We obtain the points on the transfer characteristic by first determining $i_B$ for various values of $v_s$. We then use these values to find the corresponding values of $v_{CE}$ from the load line in Fig. 3-16 in the forward-active region. In saturation, $v_{CE} < 0.3$ V.

The exact transfer characteristic in Fig. 3-29*b* is rounded near both cutoff and saturation. We can attribute this to the fact that the BJT does not turn on abruptly in the vicinity of cut in but, as seen in Fig. 3-20*b*, there is a knee in the input characteristics. Similarly, the knee in the output characteristics of Fig. 3-16 near saturation show the curves for various values of $I_B$ are bunched.

The approximate piecewise linear transfer characteristic assumes abrupt transitions from cutoff to the active region and from the active mode to saturation. As shown in Fig. 3-29*b*, this is a good approximation of the exact characteristic and, therefore, is often used.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

We note from Fig. 3-31 that the small change in $i_B$ caused by the signal ($I_{bm}$ = 20 $\mu$A) results in $I_{cm}$ = 4 mA and $V_{cem}$ = 2 V. The increased signal levels at the output are an indication of the amplification provided by the circuit.

**Notation**

At this point it is important to make a few remarks on transistor symbols. Specifically, instantaneous values of quantities which vary with time are represented by lowercase letters ($i$ for current, $v$ for voltage, and $p$ for power). Maximum, average (dc), and effective, or root-mean-square (rms), values are represented by the uppercase letter of the proper symbol ($I$, $V$, or $P$). Average (dc) values and instantaneous total values are indicated by the uppercase subscript of the proper electrode symbol ($B$ for base, $C$ for collector, $E$ for emitter). Varying components from some quiescent value are indicated by the lowercase subscript of the proper electrode symbol. A single subscript is used if the reference electrode is clearly understood. If there is any possibility of ambiguity, the conventional double-subscript notation should be used. For example, in Fig. 3-31, we show collector and base currents and voltages in the common-emitter transistor configuration, employing the notation just described. The collector and emitter current and voltage component variations from the corresponding quiescent values are

$$i_c = i_C - I_C = \Delta i_C \qquad v_c = v_C - V_C = \Delta v_C \tag{3-27}$$
$$i_b = i_B - I_B = \Delta i_B \qquad v_b = v_B - V_B = \Delta v_B$$

The magnitude of the supply voltage is indicated by repeating the electrode subscript. This notation is summarized in Table 3-4.

It is apparent from the previous paragraph that we are interested in changes in voltage and current about the operating point that occur because of the applied signal. In the next section we show that $V_{sm}$ = 26.5 mV is required to make $I_{bm}$ = 20 $\mu$A. Thus the voltage gain (or amplification) $A_V$ is

$$|A_V| = \frac{V_{cm}}{V_{sm}} = \frac{2}{26.5 \times 10^{-3}} = 75.5$$

**TABLE 3-4 Notation Summarized**

| *Notation* | *Base (collector) voltage with respect to emitter* | *Base (collector) current toward electrode from external circuit* |
|---|---|---|
| Instantaneous total value | $v_B(v_C)$ | $i_B(i_C)$ |
| Quiescent value | $V_B(V_C)$ | $I_B(I_C)$ |
| Varying component value | $v_b(v_c)$ | $i_b(i_c)$ |
| Effective value of varying component (phasor, if a sinusoid) | $V_b(V_c)$ | $I_b(I_c)$ |
| Supply voltage (magnitude) | $V_{BB}(V_{CC})$ | |

The parameter $g_m = i_c/v_\pi$, called the *transconductance*, reflects the incremental changes in $i_C$ about the operating point produced by incremental changes in the emitter-base voltage. The voltage drop $i_b r_b$ is small so that changes in the base-emitter terminal voltage can be assumed to appear across the junction. Quantitatively, $g_m$ is expressable as

$$g_m = \left.\frac{\Delta i_C}{\Delta v_{BE}}\right|_{v_{CE}=\text{const}=V_{CEQ}} = \left.\frac{\partial i_C}{\partial v_{BE}}\right|_{v_{CE}=0} \tag{3-30}$$

We believe it is important to reiterate the statements concerning the meaning of $v_{ce} = 0$ made earlier in this section. Maintaining $v_{CE}$ constant infers that no incremental change in $v_{CE}$ exists. Hence $v_{ce} = 0$, and, in Fig. 3-33, this is equivalent to short-circuiting the collector and emitter. Note that we are not physically shorting terminals $c$ and $e$ in the real transistor; this means only that the signal component of $v_{CE}$ is zero.

From Eq. (3-17), $i_C = -\alpha_F i_E$ for either an *npn* or a *pnp* transistor and Eq. (3-30) becomes[1]

$$g_m = -\alpha_F \left.\frac{\partial i_E}{\partial v_{BE}}\right|_{v_{ce}=0} \tag{3-31}$$

We wish to relate $g_m$ to the conductance of the emitter-base diode. The incremental conductance of the diode is given by Eq. (2-17) as

$$g_d = \frac{di_D}{dv_D}$$

where $i_D$ and $v_D$ are the forward current and voltage of the diode. For an *npn* transistor, $v_{BE}$ forward-biases the emitter diode and $v_{BE} = v_D$; however, $i_E$ is in the opposite direction of $i_D$ (from $n$ to $p$) so that $i_E = -i_D$. Therefore, $\partial i_E/\partial v_{BE} = -di_D/dv_D$ and

$$g_m = \alpha_F g_d \tag{3-32}$$

Equation (3-32) remains valid for a *pnp* transistor because forward-biasing the emitter junction makes $i_E = i_D$ and $v_{BE} = -v_D$.

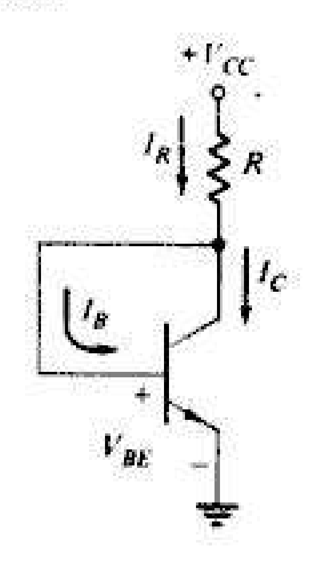
The emitter-diode conductance $g_d$ is expressed in Eq. (2-19) with $\eta = 1$. Hence $g_d = -I_{EQ}/V_T$ for an *npn* transistor and $g_d = +I_{EQ}/V_T$ for a *pnp* device. For the *npn* (*pnp*) transistor, $I_{EQ}$ is negative (positive); thus $g_d$ is positive in both instances and can be written as $g_d = |I_{EQ}|/V_T$. Using Eqs. (3-31) and (3-17) and neglecting $I_{CO}$ compared with $I_{CQ}$, we obtain the following simple expression for the transconductance:

$$g_m = \frac{\alpha_F |I_{EQ}|}{V_T} = \frac{|I_{CQ}|}{V_T} \tag{3-33}$$

---

[1]We assume that $\alpha_F$ is independent of $v_{BE}$ in Eq. (3-31).

**FIGURE 3-35**
Diode-connected transistor.




The base and collector of the transistor in Fig. 3-35 are connected; this configuration uses the emitter-base junction as the diode. The remainder of the circuit in Fig. 3-35 shows the collector (and base) returned to the supply voltage $V_{CC}$ through the resistor $R$. This positive voltage forward-biases the emitter-base junction, and as $V_{BC} = 0$ is less than the cut-in voltage, the collector-base diode is reverse-biased. Consequently, the BJT is in the forward-active region. From KVL, we obtain

$$I_R = \frac{V_{CC} - V_{BE}}{R} \tag{3-36}$$

With $V_{BE} = 0.7$ V in the active mode, the current $I_R$ is a constant and is dependent only on the supply voltage $V_{CC}$ and resistance $R$.

We can use KCL to relate $I_R$ to the transistor currents. Since $I_C = \beta_F I_B$, $I_R = I_C + I_B$ becomes

$$I_R = (\beta_F + 1)I_B = \left(1 + \frac{1}{\beta_F}\right)I_C \tag{3-37}$$

As $I_R$ is constant and with $\beta_F \gg 1$, $I_C \approx I_R$ is also constant. This observation is the basis for the current source described in the following example.

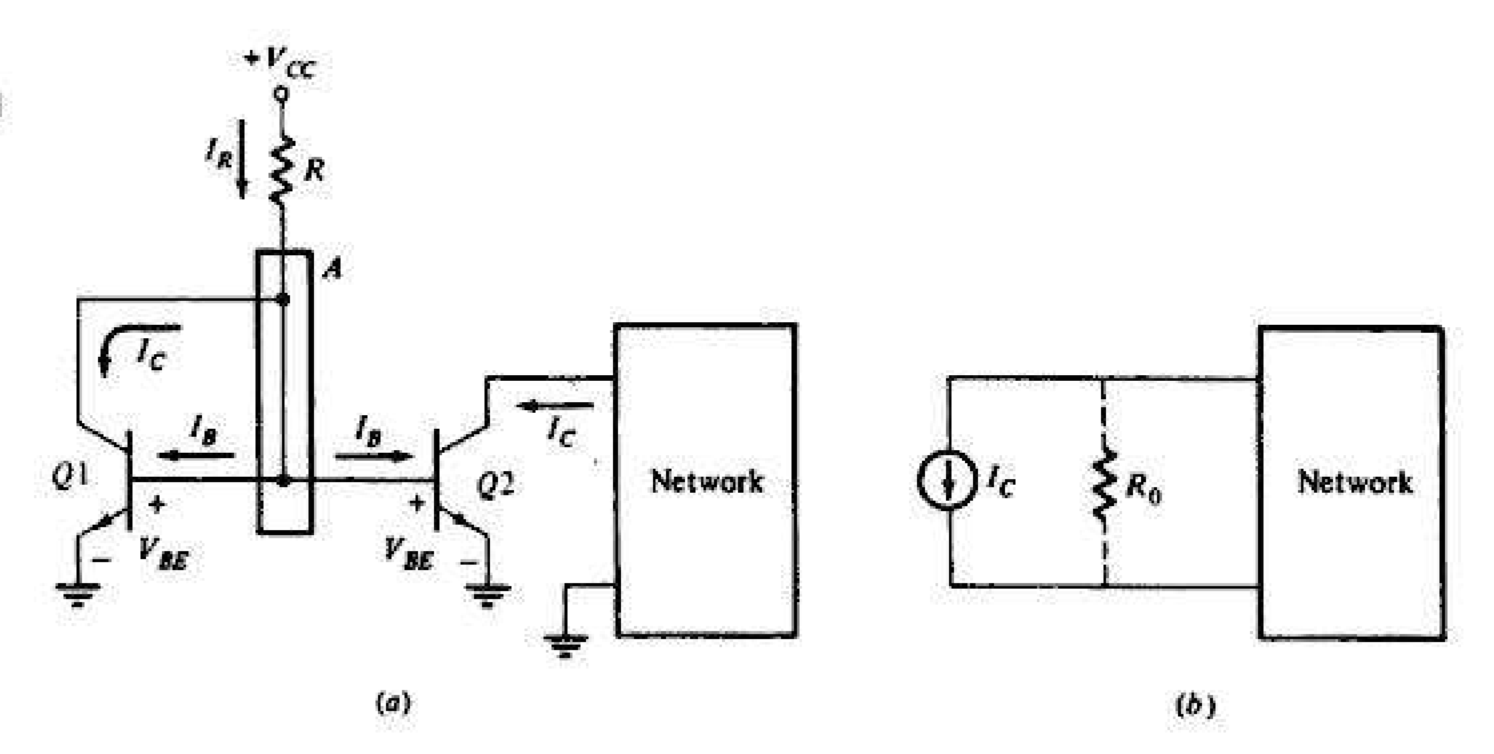
***Example 3-8***

The circuit shown in Fig. 3-36*a* is a *current source* or *current mirror* and is extensively used to bias BJTs in analog ICs. The transistors $Q1$ and $Q2$ are identical; that is, they are fabricated to have matched characteristics. (*a*) Determine $I_C$ in terms of the circuit parameters. (*b*) Evaluate $I_C$ for $V_{CC} = 10$ V, $R = 10$ kΩ, and $\beta_F = 100$. (*c*) Repeat part *b* for $\beta_F = 200$.

***Solution***

(*a*) The current $I_R$ is given by Eq. (3-36). The base-emitter voltage $V_{BE}$ of each transistor, as a consequence of KVL, is the same. Because $Q1$ and $Q2$ are identical transistors and operate at the same value of $V_{BE}$, the base and collector

**FIGURE 3-36**
(*a*) Current mirror and (*b*) its representation.



switch and that at $Q2$ as an open switch. The states of these switches are changed by applying $V_d < -4V_T$. The differential output also displays two distinct output levels, one positive and one negative, for a change in $V_d$ of about $4V_T$.

A second very important observation is that in the range $-2V_T \leq V_d \leq 2V_T$, the quantities $I_{C1}$, $I_{C2}$, $V_{o1}$, $V_{o2}$, and $V_o$ all respond in a nearly linear fashion to changes in $V_d$. For this range of inputs, the circuit behaves as a controlled source (i.e., an amplifier). Both the switching and amplifying properties of the emitter-coupled pair are employed extensively. The switching characteristic is exploited in digital circuits (Part 2), whereas the amplifying property plays a prominent role in analog circuits (Part 3).

## 3-13 TRANSISTOR RATINGS

The transistors used in the circuits described in this chapter were always assumed to be operating within allowable limits for current, voltage, and power dissipation. We now discuss the ratings (listed in the manufacturers' specification sheets) which must not be exceeded in BJT usage.

**Maximum Collector Current** Even if power and voltage ratings are not exceeded, there is an absolute maximum value of current-handling capacity of the collector, associated with the junction area and the wire bonds that connect the transistor terminals to the external leads. This rating, which determines the maximum allowable saturation current, is 800 mA for the 2N2222A transistor.

**Maximum Power Dissipation** Device destruction can occur if the collector-base junction is subjected to excessive power. The *maximum power dissipation* $P_D$ is the rating used to indicate the limits on the power-handling capability of the collector. For the 2N2222A, $P_D = 0.5$ W at an ambient temperature of 25°C. At higher ambient temperatures, $P_D$ must be derated by 12 mW/°C. Quantitatively, this means that

$$P_D(T) = 500 - 12(T - 25) \qquad \text{mW}$$

where $P_D(T)$ is the maximum power dissipation at temperature $T$ (in Celsius).

**Maximum Output Voltage Rating** There is an upper limit to the maximum allowable collector-junction voltage since there is the possibility of voltage breakdown in the transistor at high voltages. Two types of breakdown are possible, *avalanche* breakdown, discussed in Sec. 2-11, and *punch-through*, discussed in this section.

The maximum reverse-biasing voltage which may be applied before breakdown between the collector and emitter terminals of the transistor, under the condition that the base lead be open-circuited, is represented by the term $BV_{CEO}$. Breakdown may occur because of avalanche multiplication of the current $I_{CO}$ that crosses the collector junction. For 2N2222A the *CE* characteristics extending into the breakdown region are shown in Fig. 3-40, and $BV_{CEO} \approx$ 50 V. The specification sheets list the minimum value of $BV_{CEO}$ at 40 V. For

**3-25** Explain how a BJT can be used as an amplifier.

**3-26** Indicate whether each of the following is a dc, time-varying, or instantaneous quantity: $v_{ce}$, $V_{CE}$, $v_{CE}$, and $V_{ce}$.

**3-27** (*a*) Draw the hybrid-$\pi$ equivalent circuit.
(*b*) Explain the origin (or physical process) which gives rise to each term.

**3-28** Draw the low-frequency equivalent circuit of a BJT.

**3-29** (*a*) Define the transconductance $g_m$.
(*b*) Write an equation which relates $g_m$ to the current gain $\beta_o$.

**3-30** (*a*) Show by means of a circuit diagram how a BJT can be used as a diode.
(*b*) What is the region of operation of the BJT in Rev. 3-30*a*.

**3-31** (*a*) Sketch a circuit diagram of a current source.
(*b*) Briefly explain why this is a current mirror.

**3-32** Draw the circuit configuration of an emitter-coupled pair.

**3-33** Explain briefly how the differential pair can be used as an amplifier and switch.

**3-34** What limits the maximum current a transistor is capable of handling?

**3-35** (*a*) Describe punch-through.
(*b*) What limitation on transistor operation is attributable to punch-through?

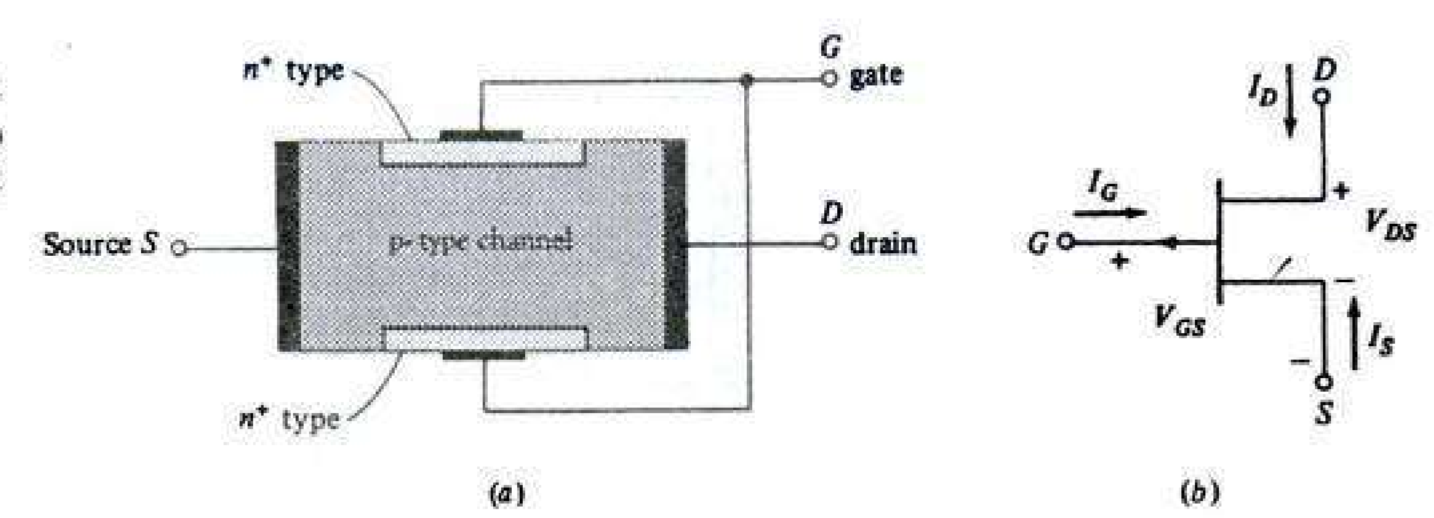



**FIGURE 4-4**
*p*-Channel junction field-effect transistor (*a*) structure and (*b*) circuit symbol.

tween the two $p^+$ gates is called the *channel* through which majority carriers move between source and drain.

The standard conventions for positive terminal currents and voltages are identified in Fig. 4-3*b*, in which the circuit symbol for an *n*-channel JFET is depicted. The corresponding structure and circuit symbol for the *p*-channel JFET are shown in Fig. 4-4.

The structures shown in Figs. 4-3 and 4-4 are convenient representations by which JFET operation can be described. The cross section of a *planar IC n-channel* JFET is shown in Fig. 4-5. The top view in Fig. 4-5 indicates how the aluminum contacts are made to the source, drain, and gate regions.

## JFET Operation

Consider the pictorial representation of an *n*-channel device displayed in Fig. 4-3 and redrawn for convenient reference in Fig. 4-6*a*. The schematic diagram in Fig. 4-6*b* corresponds to Fig. 4-6*a* and illustrates the *common-source* configuration. While our discussion focuses on an *n*-channel device, it applies equally to the *p*-channel JFET if we recognize that voltage polarities and current directions in *p*-channel devices are opposite to those for *n*-channel JFETs.

We observe that the gate regions and the channel constitute a *pn* junction which, in JFET operation, is maintained in a reverse-biased state. Application of a negative gate-to-source voltage reverse-biases the junction, as does application of a positive drain-to-source voltage. It is necessary to recall that on the two sides of the reverse-biased *pn* junction (the depletion region) there are

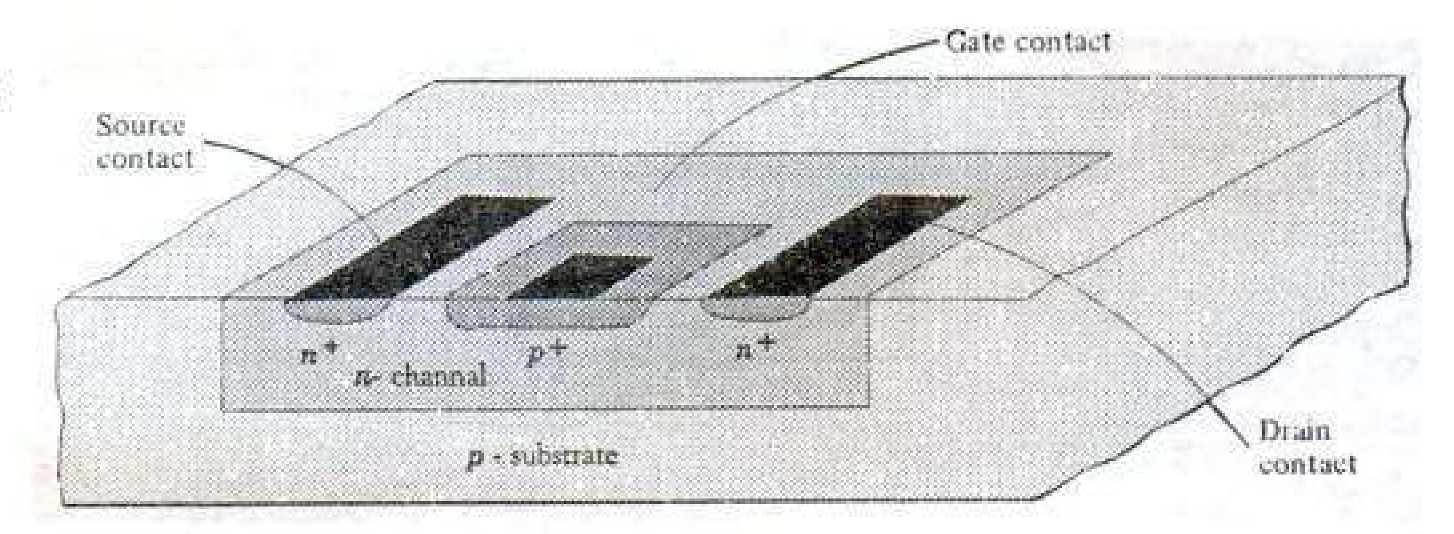



**FIGURE 4-5**
Planar IC *n*-channel JFET structure.

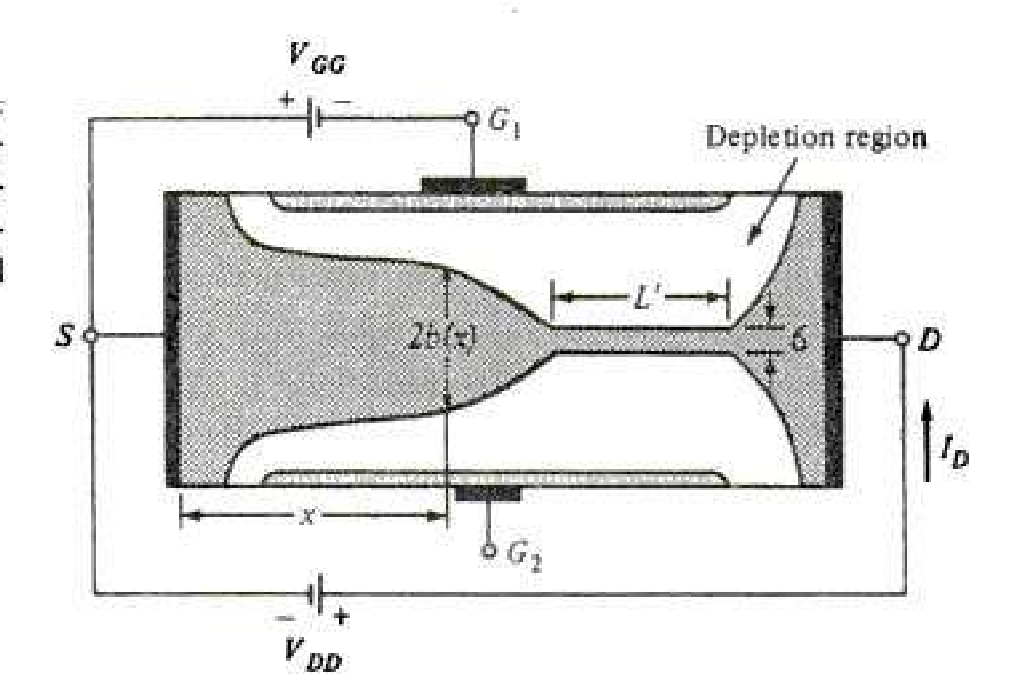



**FIGURE 4-8**
After pinch off, as $V_{DS}$ is increased, $L'$ increases, but $\delta$ and $I_D$ remain essentially constant ($G_1$ and $G_2$ are tied together).

biased than the source end, and hence the boundaries of the depletion region are not parallel to the longitudinal axis of the channel, but converge as shown in Fig. 4-8. A qualitative explanation is given in the following paragraph of what takes place within the channel as the applied drain voltage is increased and pinch-off is approached.

As $V_{DS}$ increases, $\mathscr{E}_x$ and $I_D$ increase, whereas $b(x)$ decreases because the channel narrows, and hence the current density $J = I_D/2b(x)W$ increases. We now see that complete pinch-off ($b = 0$) cannot take place, because if it did, $J$ would become infinite, which is a physically impossible condition. If $J$ were to increase without limit, it follows that [from Eq. (4-1)] $\mathscr{E}_x$ would also increase, provided $\mu_n$ remains constant. It is found experimentally, however, that the mobility is a function of electric field intensity and remains constant only for $\mathscr{E}_x < 10^3$ V/cm in $n$-type silicon. For moderate fields, $10^3$ to $10^4$ V/cm, the mobility is approximately inversely proportional to the square root of the applied field. For still higher fields, such as are encountered at pinch-off, $\mu_n$ is inversely proportional to $\mathscr{E}_x$. In this region the drift velocity of the electrons ($v_x = \mu_n \mathscr{E}_x$) remains constant, and Ohm's law is no longer valid. From Eq. (4-1) we now see that both $I_D$ and $b$ remain constant, thus explaining the constant-current portion of the $V$-$I$ characteristic illustrated in Fig. 4-7.

What happens if $V_{DS}$ is increased beyond pinch-off, with $V_{GS}$ held constant? As explained above, the minimum channel width $b_{\min} = \delta$ has a nonzero constant value. This minimum width occurs at the drain end of the bar. As $V_{DS}$ is increased, this increment in potential causes an increase in $\mathscr{E}_x$ in an adjacent channel section toward the source. Referring to Fig. 4-8, the velocity-limited region $L'$ increases with $V_{DS}$, whereas $\delta$ remains at a fixed value.

The student must take care not to confuse the different meanings of pinch-off and saturation used in describing semiconductor devices. Along the constant-current portion of the characteristic, pinch-off refers to the fact that $V_{DS}$ is used to constrict the channel almost entirely. The pinch-off voltage $V_p$ refers to the voltage applied to the gate which totally blocks the channel independent of $V_{DS}$. Saturation in an FET refers to the limiting value of drift velocity. Thus

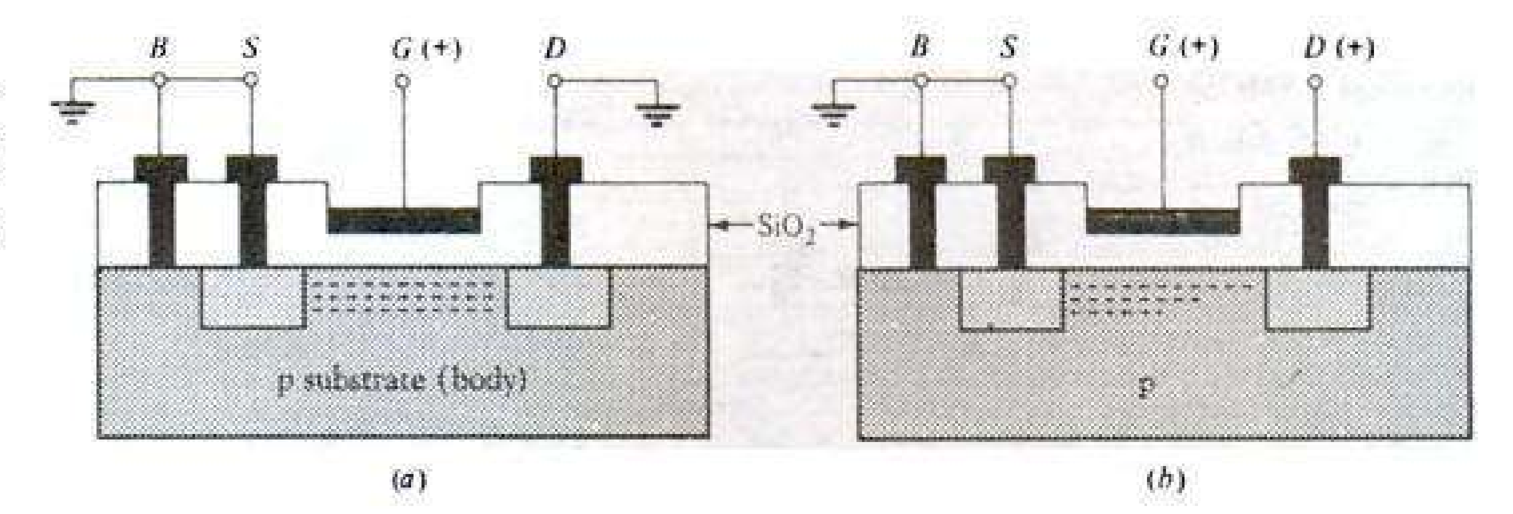
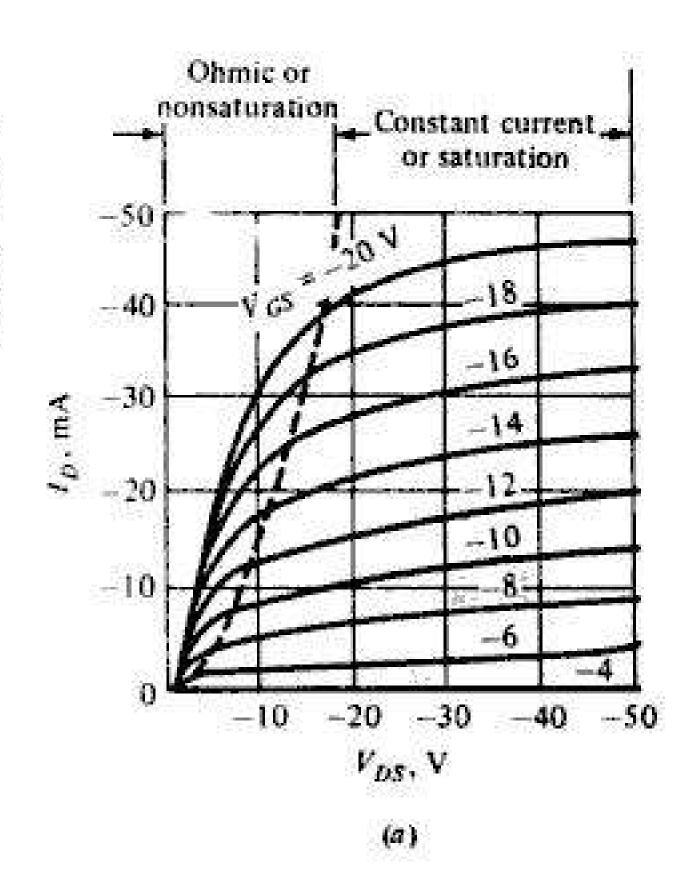



**FIGURE 4-11** Biased NMOS enhancement transistor showing induced channel with (*a*) $V_{DS} = 0$ and (*b*) $V_{DS} > 0$.

an electric field which is directed perpendicularly through the oxide. This field will end on "induced" negative charges near the semiconductor surface, as shown in Fig. 4-11*a*. Since the *p*-type substrate contains very few electrons, the positive surface charges are primarily electrons obtained from the *n*-type source and drain. These mobile negative charges, which are minority carriers in the *p*-type substrate, form an "inversion layer." Such an inversion layer is formed only if $V_{GS}$ exceeds a threshold level $V_T$.[1] The induced charges beneath the oxide constitute an *n* channel. As the voltage on the gate increases beyond $V_T$, the number of induced negative charges in the semiconductor increases. Consequently, the conductivity of the channel increases. Application of a positive potential between drain and source produces a current in the induced channel between drain and source. Thus, the drain current is enhanced by the positive gate voltage and the device is called an *enhancement-type MOSFET*.

Let us now consider the situation where $V_{DS}$ is increased from zero with $V_{GS}$ maintained at a constant positive value greater than $V_T$ (that is, $V_{GS} - V_T > 0$). For small values of $V_{DS}$ ($V_{DS} < V_{GS} - V_T$), an increase in $V_{DS}$ is accompanied by an increase in drain current $I_D$. The behavior of the MOSFET is that of a resistance, and this region is referred to as the *ohmic region*. As $V_{DS}$ increases, the drop across the channel increases in magnitude, and hence the voltage across the gate oxide at the drain side of the channel $V_{GD} = V_{DS} - V_{GS}$ decreases. This reduced potential difference lowers the field across the drain end of the dielectric, which results in fewer inversion charges in this portion of the induced channel. The channel is being "pinched off," and $I_D$ increases much more slowly with respect to increases in $V_{DS}$ than in the ohmic region near the origin. Ideally, once pinch-off is achieved, a further increase in $V_{DS}$ produces no change in $I_D$ and current saturation exists. This *saturation region* is similar in nature to velocity saturation in a JFET. The value of $I_D$ attained in saturation depends on the value of $V_{GS}$. Increases in $V_{GS} > V_T$ result in increasing saturation values of $I_D$.

[1] In this chapter the threshold voltage should not be confused with the volt-equivalent of temperature described in Sec. 1-6.



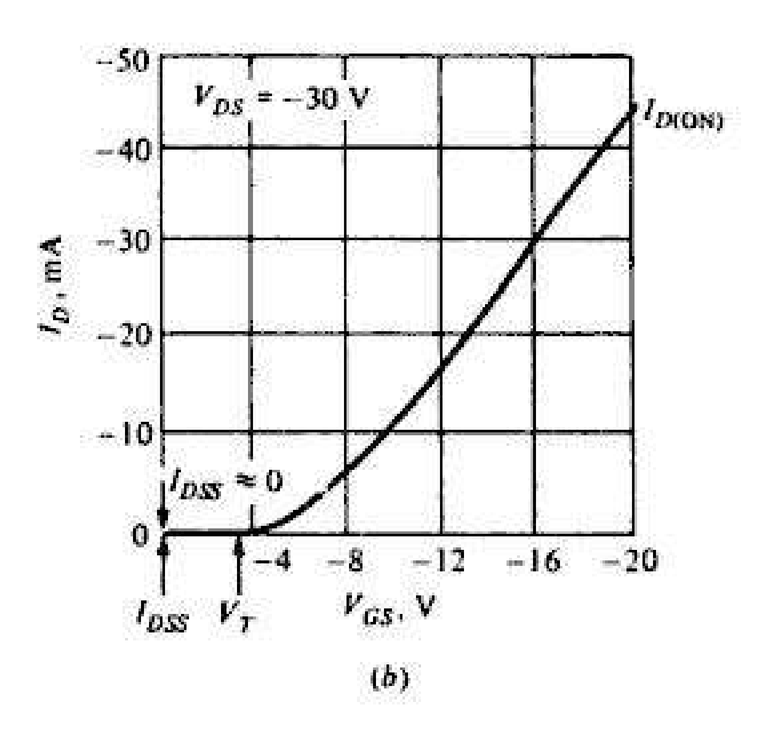




**FIGURE 4-15**
(*a*) The drain characteristics and (*b*) the transfer characteristic for $V_{DS} = -30$ V of the 3N133 enhancement PMOS. (Courtesy of Siliconix, Inc.)

**$p$-Channel Enhancement MOSFET Characteristics** The PMOS transistor, depicted in Fig. 4-10*b*, consists of two $p$-type regions in an $n$-type substrate. The physical principles of a $p$-channel enhancement MOSFET are the same as in NMOS devices. Equations (4-4) to (4-7) are applicable once we recognize that all voltage polarities and current directions in a PMOS device are opposite to the corresponding quantities in NMOS transistors. Also, in evaluating the process parameter $k$, $\mu_n$ is replaced by the hole mobility $\mu_p$.

The volt-ampere characteristics of a discrete PMOS transistor is shown in Fig. 4-15*a*; its transfer characteristic is depicted in Fig. 4-15*b*. Note that the general shape of these curves is similar to the NMOS characteristics in Figs. 4-12 and 4-13. However, observe the differences in current and voltage levels for these devices. In discrete MOSFET fabrication, device dimensions can be made larger than are convenient in IC MOSFETs. Consequently, discrete MOS transistors which have higher current capabilities and which can operate with larger applied voltages are obtainable.

**Comparison of PMOS and NMOS Transistors** Historically, $p$-channel enhancement transistors were used first in MOS systems because they were more easily produced with greater yields and reliability than $n$-channel devices. Improvements in fabrication methods have led to the dominance of NMOS transistors, and—with the exception of CMOS technology (Sec. 4-15)—have made PMOS devices almost obsolete. The reasons for this are described in the next paragraph.

The hole mobility in silicon at normal field intensities is about 500 cm²/(V·s). On the other hand, electron mobility is about 1300 cm²/(V·s). Thus, for devices having the same dimensions (1) the current in a PMOS transistor is less than half of that in an NMOS device and (2) the ON resistance of a

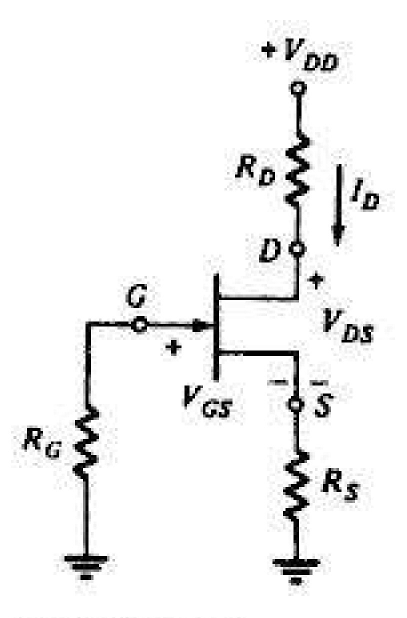



**FIGURE 4-19**
Self-biased JFET stage.

$$V_{GS} = -I_D R_S \quad \text{or} \quad I_D = \frac{-V_{GS}}{R_S} \tag{4-8}$$

Equation (4-8) defines a straight line called the *bias line* and is plotted on the JFET transfer characteristic shown in Fig. 4-20. The intersection of the transfer characteristic and the bias line determines the operating (quiescent) values of drain current $I_{DQ}$ and gate-to-source voltage $V_{GSQ}$.

The drain-to-source voltage $V_{DSQ}$ is evaluated from the KVL equation for the drain-source loop. This expression is

$$-V_{DD} + I_D R_D + V_{DS} + I_D R_S = 0 \tag{4-9}$$

Substitution of $I_{DQ}$ into Eq. (4-9) gives the quiescent value of drain-to-source voltage $V_{DSQ}$ that exists in the circuit. Note that Eq. (4-9) specifies the load line for the circuit. By constructing the load line on the output characteristics, we can also determine the value of $V_{DSQ}$ from the intersection of the load line with the characteristic for $V_{GSQ}$.

*Example 4-1*

Consider the circuit in Fig. 4-21*a*, which employs an *n*-channel enhancement MOSFET whose transfer characteristic is shown in Fig. 4-13 and is repeated for convenience in Fig. 4-22. Determine $I_{DQ}$, $V_{DSQ}$, and the output quiescent voltage $V_{OQ}$.

*Solution*

First replace the gate-bias resistances $R_1$ and $R_2$ and the drain supply voltage $V_{DD}$ by their Thévenin equivalents as shown in Fig. 4-21*b*. (Note that this is analogous to the analysis of the BJT circuit in Example 3-4.)

The bias-line equation is obtained from the KVL expression for the gate-source loop in Fig. 4-21*b* and is

$$V_{GS} = V_{GG} - I_D R_S \quad \text{or} \quad I_D = -\frac{1}{R_S} V_{GS} + \frac{V_{GG}}{R_S}$$

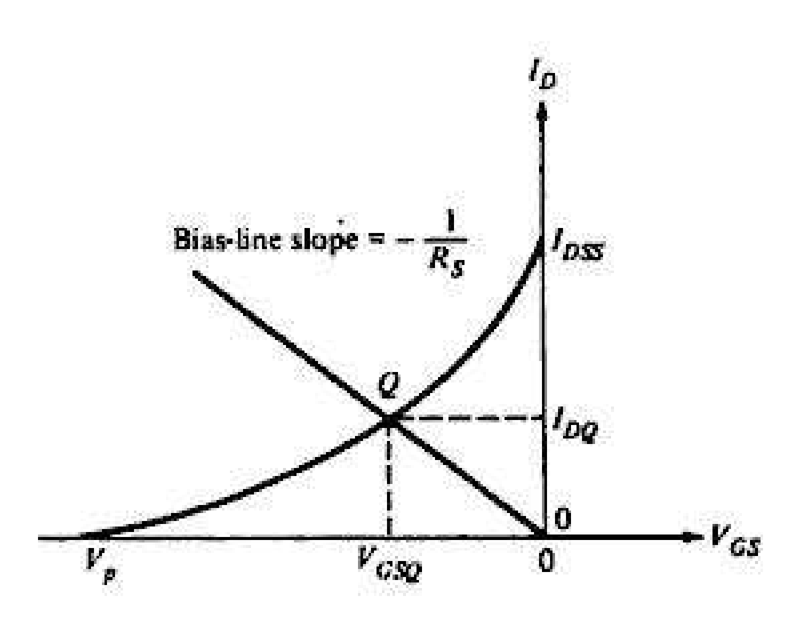



**FIGURE 4-20**
The bias line, determined by $R_S$, is drawn on the transfer characteristic. The intersection $Q$ is the quiescent point, and the drain current and the gate-to-source voltage that exist in the circuit are indicated by $I_{DQ}$ and $V_{GSQ}$, respectively.

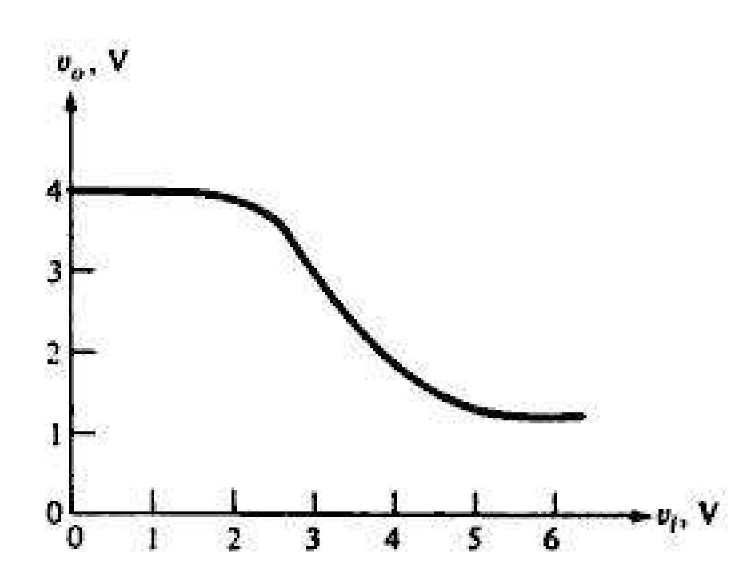



**FIGURE 4-25**
Voltage transfer characteristic ($v_o$ versus $v_I$) for Example 4-2.

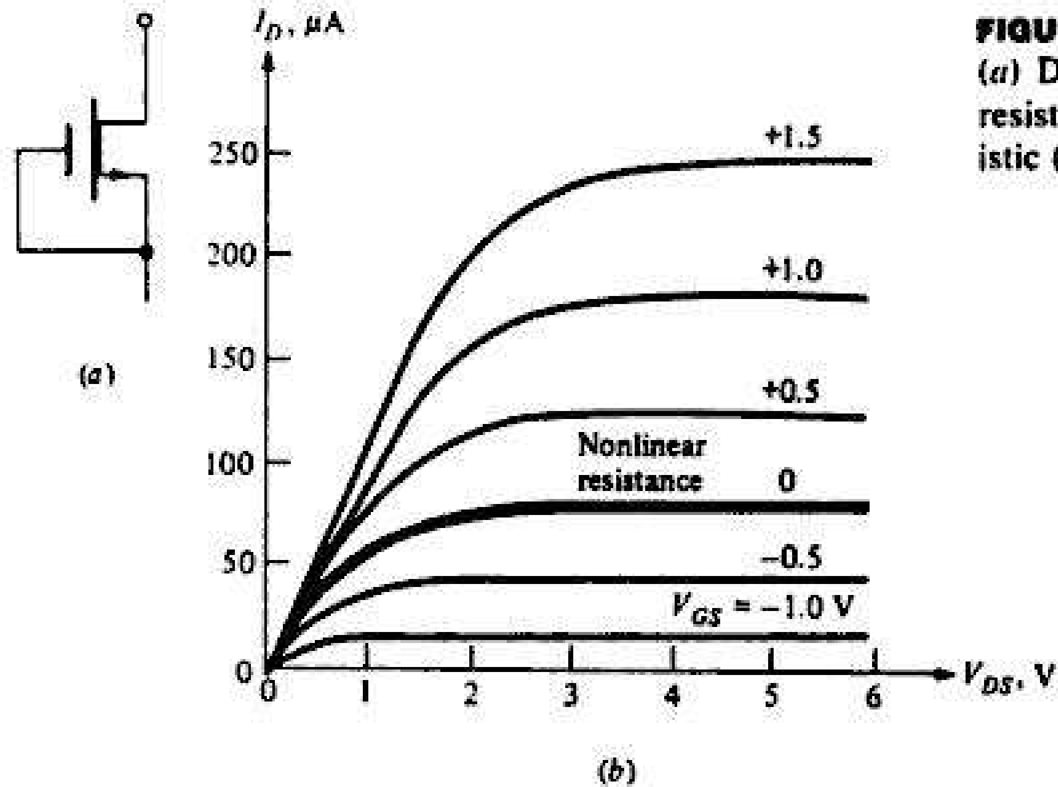



**FIGURE 4-26**
(*a*) Depletion-mode MOSFET connected as a resistance; (*b*) nonlinear resistance characteristic ($V_{GS}$ = 0).

**4-12 THE FET AS A SWITCH** Metal oxide semiconductor field-effect transistors are used extensively in digital circuits which exploit the behavior of these devices as switches. We demonstrate the controlled-switch operation of the MOSFET using the circuit shown in Fig. 4-24*a* and Example 4-2. The input voltage waveform is the step displayed in Fig. 4-27*a*. For $t < T$, the input voltage is 1.5 V; thus, from the transfer characteristic in Fig. 4-25, we observe that $v_o$ = 4 V. The current in the circuit $I_{D1}$ is zero as determined from the load line in Fig. 4-24*b*. This is the characteristic of an open switch as the voltage across the switch is "appreciable" while the current is zero.

For $t > T$, the input voltage is 5.0 V, resulting in $v_o$ = 1.5 V (Fig. 4-25) and $I_{D1}$ = 125 μA (Fig. 4-24*b*). This condition approximates a closed switch as "appreciable" current exists with a low voltage across the switch terminals (*D* and *S*). The output waveform is displayed in Fig. 4-27*b*.

The two states of the switch can be discerned from the transfer characteristic given in Fig. 4.25. As long as $v_I \leq V_T$ = 2 V, the output voltage is 4 V and,

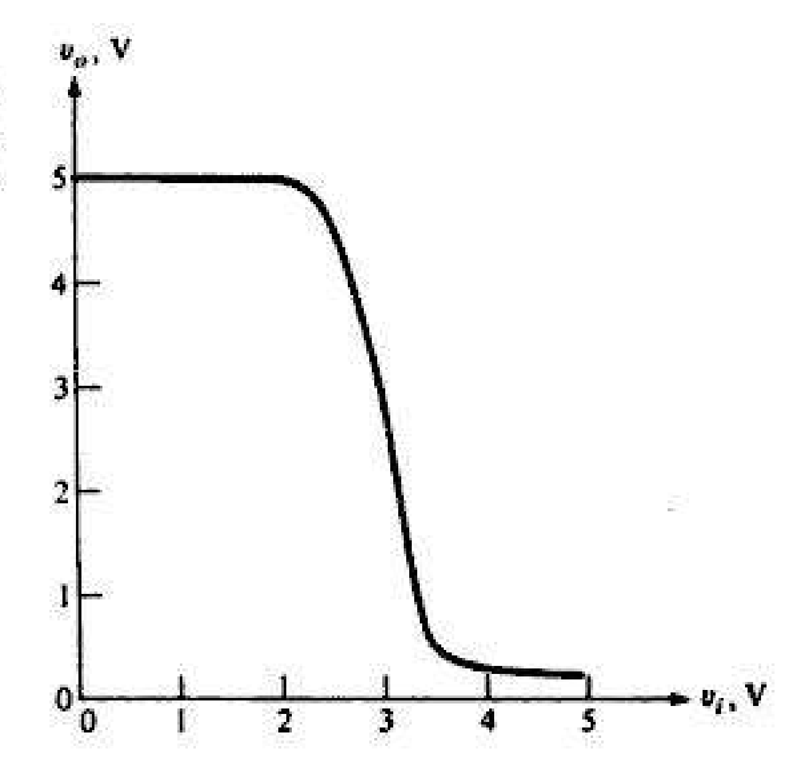



**FIGURE 4-30**
Voltage transfer characteristic of circuit in Fig. 4-29*a* and Example 4-4.

## 4-13 THE FET AS AN AMPLIFIER

Field-effect transistor amplifier circuits exploit the voltage-controlled current-source nature of these devices. As discussed earlier in this chapter, it is in the saturation region where $I_D$ depends (approximately) only on $V_{GS}$, thus making available controlled-source behavior. We demonstrate the use of the FET as an amplifier by considering the common-source circuit in Fig. 4-31.

The signal to be amplified in Fig. 4-31 is $v_s$, whereas $V_{GG}$ provides the necessary reverse-bias between gate and source of the JFET. The volt-ampere characteristics of the JFET are shown in Fig. 4-32 upon which a load line corresponding to $V_{DD} = 30$ V and $R_D = 6$ kΩ is constructed. The value of $V_{GG}$ is selected as 1.5 V so that the transistor is biased at point $Q$ and results in $V_{DSQ} = 19$ V and $I_{DQ} = 1.8$ mA.

The instantaneous gate-to-source voltage is $v_{GS} = v_s - V_{GG}$. Assuming that $v_s$ is a sinusoid of peak voltage $V_m = 0.5$ V, the variation with time in $v_{GS}$, shown in Fig. 4-32, is a sinusoid superimposed on the quiescent level. The resultant waveforms for $i_D$ and $v_{DS}$ are displayed alongside the characteristics.

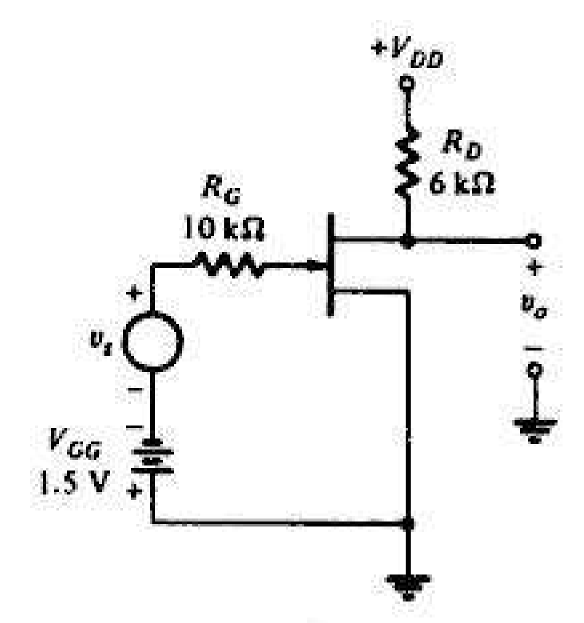



**FIGURE 4-31**
Common-source JFET amplifier stage.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

Since we have demonstrated that $g_m$ is always positive, this equation can be written in the alternative form

$$g_m = -\frac{2I_{DSS}}{V_p}\sqrt{\frac{I_{DQ}}{I_{DSS}}} = g_{mo}\sqrt{\frac{I_{DQ}}{I_{DSS}}} \tag{4-18}$$

The term $g_{mo} = 2I_{DSS}/V_p$ is the value of $g_m$ when $V_{GSQ} = 0$ fór which $I_{DQ} = I_{DSS}$.

Similarly, for an NMOS transistor, $g_m$ is expressable as

$$g_m = 2\sqrt{k\left(\frac{W}{L}\right) I_{DQ}} \tag{4-19}$$

[The derivation of Eq. (4-19) is left to the student as an exercise in Prob. 4-39.]

Because $r_{ds}$ reflects the effect of channel-length modulation, Eq. (4-7) is used to relate $i_D$ and $v_{DS}$. For a MOSFET, the output conductance $g_{ds}$ is expressed as

$$g_{ds} \equiv \frac{1}{r_{ds}} = \left.\frac{\partial i_D}{\partial v_{DS}}\right|_{v_{GS}=V_{GSQ}} = \left.\frac{i_d}{v_{ds}}\right|_{v_{gs}=0} \tag{4-20}$$

Application of Eq. (4-20) in conjunction with Eq. (4-7) yields

$$g_{ds} = \lambda k\left(\frac{W}{L}\right)(V_{GSQ} - V_T)^2 = \frac{\lambda I_{DQ}}{1 + \lambda V_{DSQ}} \tag{4-21}$$

Hence

$$r_{ds} = \frac{1 + \lambda V_{DSQ}}{\lambda I_{DQ}} \tag{4-22}$$

For IC FETs, Eq. (4-22) is usually evaluated at $V_{DSQ} = 0$ and reduces to

$$r_{ds} = \frac{1}{\lambda I_{DQ}} \tag{4-23}$$

Normal operation of IC FETs is at drain-to-source voltages in the order of a few volts. Consequently, the term $\lambda V_{DSQ}$ in Eq. (4-22) is much less than unity so that Eq. (4-23) is a good approximation of Eq. (4-22). For discrete FETs, particularly those which are used at moderate voltage and power levels, $r_{ds}$ is evaluated by using Eq. (4-22). (See Figs. 4-7 and 4-32 for the 2N4869 $n$-channel JFET used in Example 4-5.)

Equation (4-22) is also valid for a JFET as the channel-length modulation term $(1 + \lambda v_{DS})$ can also be incorporated into Eq. (4-3).

***Example 4-5*** Determine the voltage gain of the JFET amplifier stage shown in Fig. 4-31. The JFET has $I_{DSS} = 5$ mA, $V_p = -3.6$ V, and $\lambda = 0.01$ V$^{-1}$.

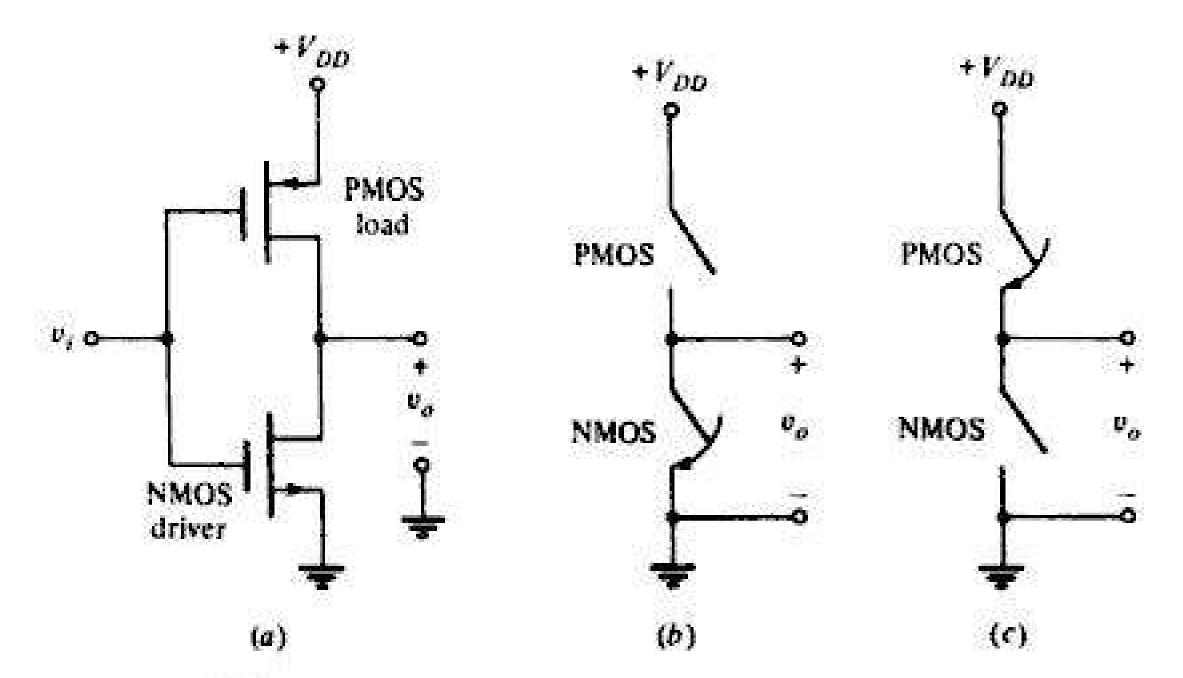



**FIGURE 4-38**
(*a*) Complementary metal oxide semiconductor switch containing NMOS driver and PMOS load. Ideal switch representation of CMOS circuit in part *a* when NMOS transistor shown in part *b* is ON and (*c*) when PMOS transistor conducts.

The CMOS circuit in Fig. 4-38*a*, employed in digital applications, consists of an NMOS transistor (the driver) to which a PMOS transistor is connected as the load. The gate terminals of the two transistors are connected to one another. Let us consider that the magnitude of the threshold voltage $V_T$ of each transistor is the same and equals $V_{DD}/2$. Application of a positive value of $v_i > V_T$ simultaneously turns on the NMOS transistor and cuts off the PMOS device. (Recall that a positive gate voltage is required in *n*-channel transistors and negative gate voltages are needed in *p*-channel devices.) Because the drain and source terminals of both transistors are series-connected, no current exists in the NMOS device (as a result of the cutoff condition in the PMOS transistor). Consequently, the output voltage is virtually zero. This situation is idealized in Fig. 4-38*b* in which the closed switch represents the NMOS devise and the open switch, the PMOS device.

Similarly, when a negative (or zero) input voltage is applied, the PMOS load is turned on, but the NMOS driver is cut off. The switch conditions in Fig. 4-38*c* approximate this situation. Again, because one transistor is cut off, no current exists in the circuit. The output voltage is high; it is $V_{DD}$ in the idealized case shown in Fig. 4-38*c* and is approximately $V_{DD}$ in Fig. 4-38*a*.

The action described is that of opening and closing a switch by means of the input control voltage. However, because no current exists in either switch state, the power dissipated by the transistors is virtually zero. (Indeed, power is consumed in the CMOS device only during the switching interval.) The extremely small power consumption of CMOS circuits is the major motivation for their widespread use.

Complementary metal oxide semiconductor analog circuits often employ the configuration depicted in Fig. 4-39*a*. The PMOS transistor provides the resistive load for the NMOS transistor, which functions as the controlled source.

# Chapter 5
# INTEGRATED-CIRCUIT FABRICATION

An integrated circuit (IC) consists of a single crystal chip of silicon, typically 0.25 mm thick and covering a surface area 1 to 10 mm by 1 to 10 mm containing both active and passive elements. The processes used to fabricate ICs are described qualitatively in this chapter. These processes include wafer preparation, epitaxial growth, impurity diffusion, ion implantation, oxide growth, photolithography, chemical etching, and metallization. Batch processing is employed and offers excellent repeatability for the production of large numbers of ICs at low cost.

Each step in the fabrication contributes to both the capabilities and limitations of the circuits produced. The object of this discussion is to present an overview of IC technology and the implications of this technology on circuit design. Specifically, both bipolar and MOS fabrication are treated.

## 5-1 MONOLITHIC INTEGRATED-CIRCUIT (MICROELECTRONIC) TECHNOLOGY

The term "monolithic" is derived from the Greek words *monos* (meaning "single") and *lithos* (meaning "stone"). Thus a monolithic integrated circuit is built into a single "stone" or single crystal of silicon. The word "integrated" refers to the fact that all the circuit components—transistors, diodes, resistors, capacitors, and their interconnections—are fabricated as a single entity. Note that inductors are not included; one of the consequences of semiconductor IC construction is that practical values of inductance cannot be realized.[1]

The variety of manufacturing processes by which ICs are fabricated take place through a single plane and hence can be conveniently called *planar technology*. The structure of a simple bipolar IC is shown in Fig. 5-1*a* and is the

[1]In Parts 3 and 4 we introduce some circuits whose performance characteristics are similar (or identical) to those containing inductors.

The following advantages are offered by IC technology as compared with discrete components interconnected by conventional techniques:

1. Low cost (due to large quantities processed).
2. Small size.
3. High reliability. All components are fabricated simultaneously, and with no soldered joints, both mechanical and electrical failure are reduced.
4. Improved performance. Because of the low cost, more complex circuitry may be used to obtain better functional characteristics.
5. Matched devices. Since all transistors are manufactured simultaneously by the same processes, the corresponding parameters of these devices as well as the temperature variation of their characteristics have essentially the same magnitudes (the parameters track well with temperature).

## 5-2 THE PLANAR PROCESSES

As briefly introduced in Sec. 5-1, planar technology for IC fabrication consists of six or seven independent processes: (1) crystal growth of the wafer, (2) epitaxial growth, (3) oxidation, (4) photolithography and chemical etching, (5) diffusion, (6) ion implantation, and (7) metallization. We now examine each of these processes in more detail.

### Crystal Growth of the Substrate

A tiny seed crystal of silicon is attached to a rod and lowered into a crucible of molten silicon to which acceptor impurities have been added. As the rod is very slowly pulled out of the melt under carefully controlled conditions, a single *p*-type crystal ingot of the order of 4 in (10 cm) in diameter and 20 in (50 cm) long is grown. This technique is referred to as the *Czochralski* or *CZ* process. The ingot is subsequently sliced into round wafers approximately 0.2 mm thick to form the substrate on which all integrated components will be fabricated. One side of each wafer is lapped and polished to eliminate surface imperfections before proceeding with the next process.

### Epitaxial Growth

The word "epitaxy" comes from the Greek words *epis*, meaning layered and *taxis*, meaning ordered. In IC fabrication, the epitaxial process is used to grow a layer of single-crystal silicon as an extension of an existing crystal wafer of the same material. Epitaxial growth is performed in a special furnace called a *reactor* into which the finished silicon wafers are inserted and heated to 900 to 1000°C. Current production technology uses the hydrogen reduction of the gases silane ($SiH_4$) or silicon tetrachloride ($SiCl_4$) as the source of the silicon to be grown. Silane has the advantages of requiring lower temperature and having a faster growth rate than does $SiCl_4$.

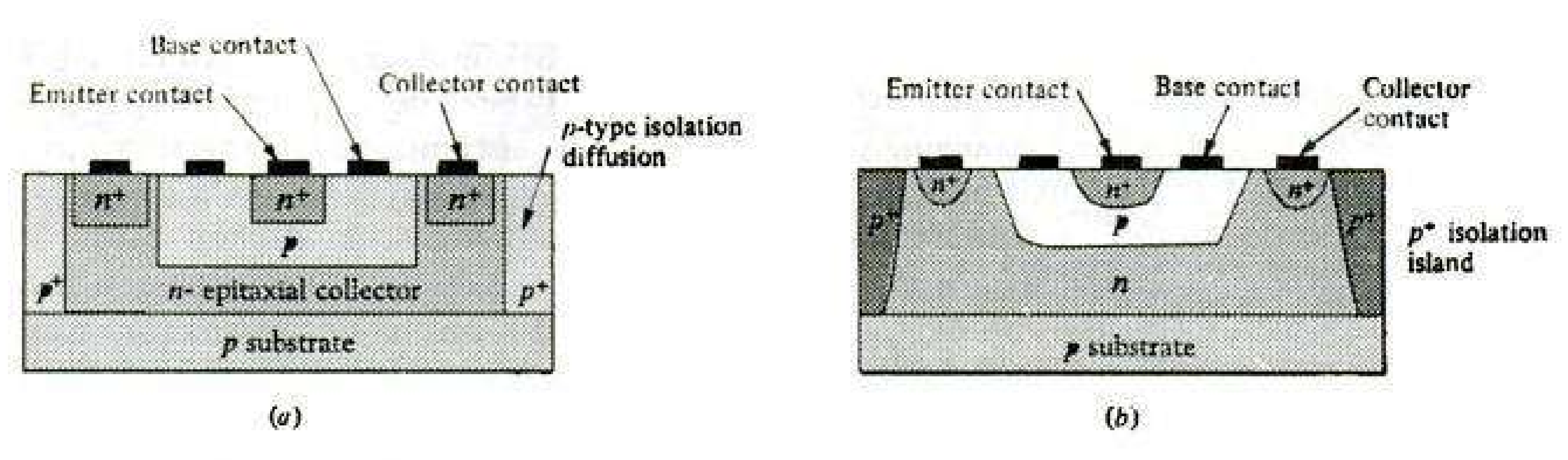


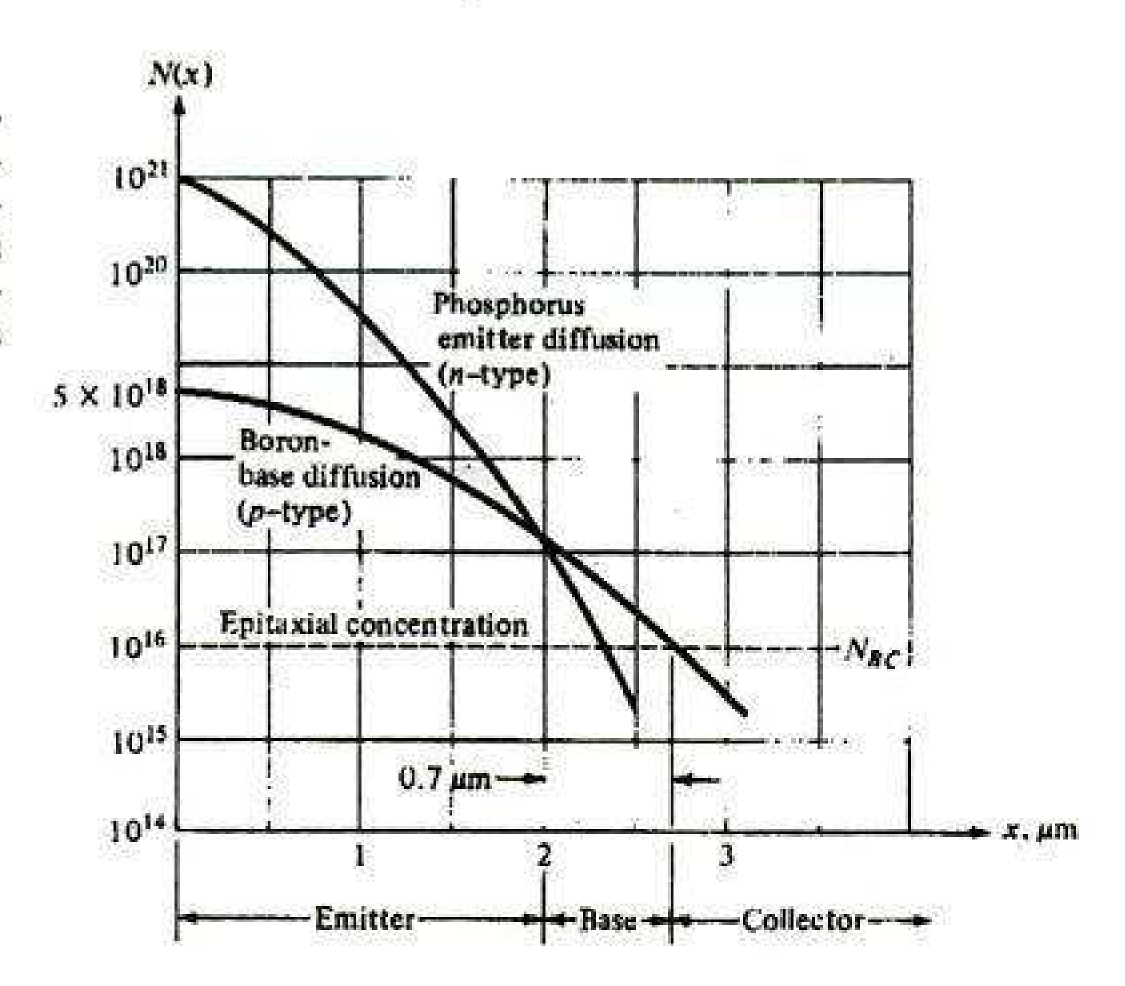

**FIGURE 5-5**
Cross sections of an IC transistor: (a) idealized and (b) actual.

the same distance that they do vertically. Hence the impurity will spread out under the passivating oxide surface layer, and the junction profiles should be drawn more realistically as shown in Fig. 5-5b.

In a bipolar transistor, two diffusions of impurities are often used. For an *npn* device, the first is the diffusion of the *p*-type base into the *n*-type epitaxially grown collector; the second is the *n*-type emitter region into the *p*-type base. A typical impurity profile for a double-diffused *npn* monolithic transistor is displayed in Fig. 5-6.

The background, or epitaxial-collector, concentration $N_{BC}$ is represented by the dashed line in Fig. 5-6. The concentration $N$ of boron is high ($5 \times 10^{18}$ atoms/cm$^3$) at the surface and falls off with distance into the silicon as indicated

**FIGURE 5-6**
A typical impurity profile in a monolithic double-diffused planar transistor. Note that $N(x)$ (in atoms per cubic centimeter) is plotted on a logarithmic scale.



BJTs having higher current capabilities (increasing $I_{ES}$), the emitter areas are increased. Consequently, the entire device area is also increased. A rule of thumb often used is to restrict emitter-area ratios to 10:1 for transistors in close proximity. This is due to the fact that chip area is at a premium and to the limitations of the diffusion process.

Commercial IC processes often employ ion implantation of the emitter and base areas. These regions are shallow, and their depths can be more accurately controlled by implantation. Furthermore, as ion implantation is performed at lower temperatures than diffusion, the disadvantage of lateral spreading of the base and emitter is minimized.

**Buried Layer**

The fabrication of the BJTs indicated in Fig. 5-7 is almost always modified with an additional processing step, as identified in Fig. 5-8 (and shown in Fig. 5-7). The two $n^+$ regions, referred to as the *buried layer*, at the interface between the $n$ and $p$ layers is deposited prior to epitaxial growth. Recall that the $n^+$ designation indicates an $n$ region of higher dopant concentration than one simply identified as $n$ type. The use of the $n^+$ regions has two functions: (1) it enhances the growth of the epitaxial layer, and (2) the higher electron density in the $n^+$ area reduces the series resistance between the collector junction and the collector terminal (see Sec. 3-7).

***pnp* Fabrication**

The vast majority of IC bipolar transistors are *npn*. However, in some circuits, *pnp* devices are required. For example, in the emitter-coupled pair described in Sec. 3-12, the collector resistors are generally realized by using a pair of *pnp* transistors in a current-source configuration. The *lateral pnp* and *vertical* or *substrate pnp* are the two kinds of such transistors usually employed.

In Fig. 5-8 we see that the base, collector, and isolation regions form a parasitic *pnp* transistor. The term "lateral" refers to the fact that the three elements lie in a horizontal plane in contrast to the vertical plane of the *npn* transistors. Similarly, a parasitic vertical *pnp* device is formed by the base and collector of the *npn* transistor and the *p*-type substrate. These observations lead to the construction of the two types of *pnp* transistors employed in bipolar ICs.

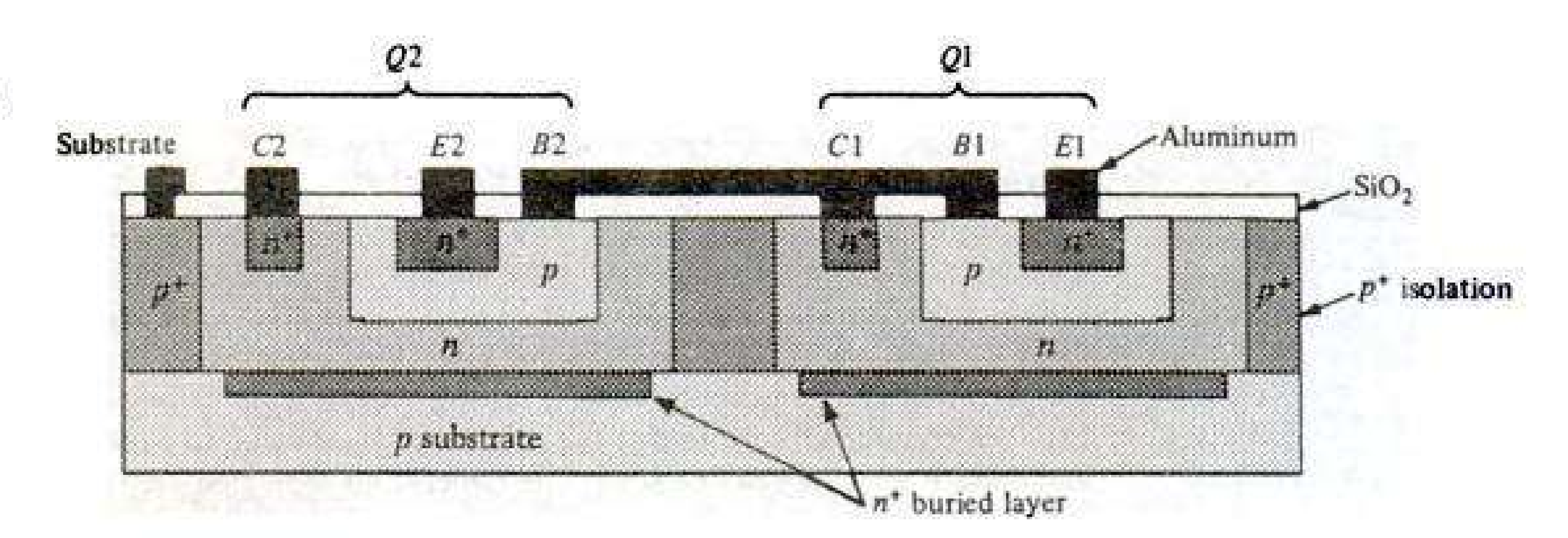



**FIGURE 5-8**
IC transistors showing buried layer.

**5-4 FABRICATION OF FETs** The fabrication of MOS transistors is demonstrated by describing the process sequence used to construct the NMOS enhancement and depletion devices as shown in Fig. 5-2. In this section, JFET construction is also included.

**NMOS Enhancement Fabrication** The first step in the sequence is to coat the entire surface of a *p*-type wafer with silicon nitride ($Si_3N_4$). In Sec. 5-2 we noted that $Si_3N_4$ is more impervious to the dopants used than is $SiO_2$. The first mask and etching step is used to define an area large enough to include the source, gate, and drain. The $Si_3N_4$ is chemically removed from the surface outside of the transistor region. Next, a $p^+$ layer is ion-implanted near the surface of the exposed *p* substrate. The $p^+$ implant serves to isolate adjacent devices as explained in Sec. 5-3. This part of the processing sequence is completed by growing a thick (1-$\mu$m) $SiO_2$ layer, called the *field oxide*, over the $p^+$-implanted regions as illustrated in Fig. 5-14*a*. (The $Si_3N_4$ region is unaffected by oxidation.)

In the second phase of processing, the remaining $Si_3N_4$ (but not the $SiO_2$) is removed by selective etching and a thin $SiO_2$ layer (800 to 1000 A) is thermally grown over the transistor areas (Fig. 5-14*b*). This process provides the oxide layer that lies under the gates of the transistors.

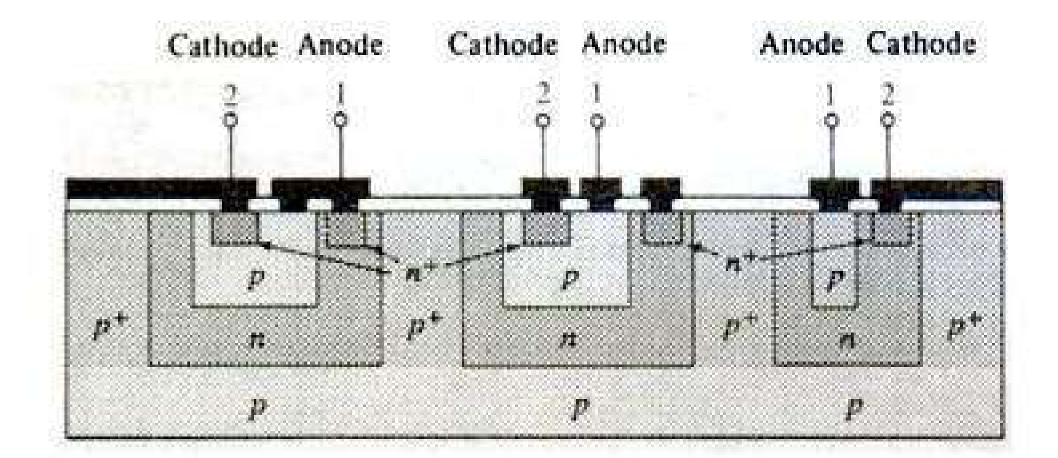
Polycrystalline silicon, colloquially referred to as *polysilicon* or simply *poly*, is now deposited over the entire wafer. A second photolithographic process defines the gate region and the excess polysilicon is removed by etching. The resulting cross section and top view of the chip are depicted in Figs. 5-14*c* and 5-14*d*. The use of polysilicon gates reduces the threshold voltage $V_T$ below the values obtainable with metal gates. As a result, lower supply voltages can be used with polysilicon gates. Consequently, most commercial MOS ICs are fabricated with polysilicon gates.

The $n^+$ source and drain regions are usually obtained by ion implantation. The field oxide and polysilicon gate prevent the penetration of dopants below these regions. The thin oxide layer, however, is penetrated by the dopant and allows the formation of drain and source. As a result of ion implantation, there is a *self-alignment* of both gate and source and gate and drain. Since there is no overlap of these electrodes, the capacitances between them, $C_{gs}$ and $C_{gd}$, are drastically reduced.

After the implantation of source and drain, the entire wafer is covered with a protective insulating layer (usually $SiO_2$). A third mask is used to define the contacts to the device (including the body *B*) and etching exposes the contact surfaces. Aluminum is evaporated over the entire wafer and a fourth mask is used to pattern the circuit connections (Figs. 5-14*e* and 5-14*f*). Note that Fig. 5-14*e* corresponds to *Q*1 in Fig. 5-1*b*.

**Self-Isolation** The $p^+$ implant in Figs. 5-14*e* and 5-14*f* serves as a low resistance for the contact *B* to the substrate (body) of the MOSFET. Usually the source and body are tied together, as in Fig. 5-14*e*, and thus the source-substrate diode is cut off. The drain potential polarity for an NMOS device is positive with respect

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.



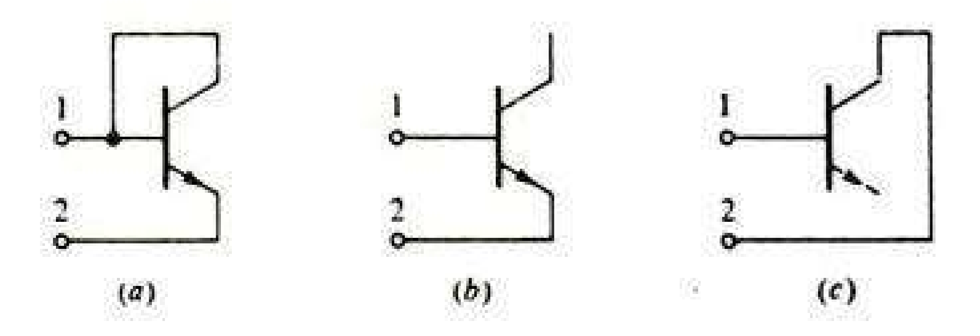
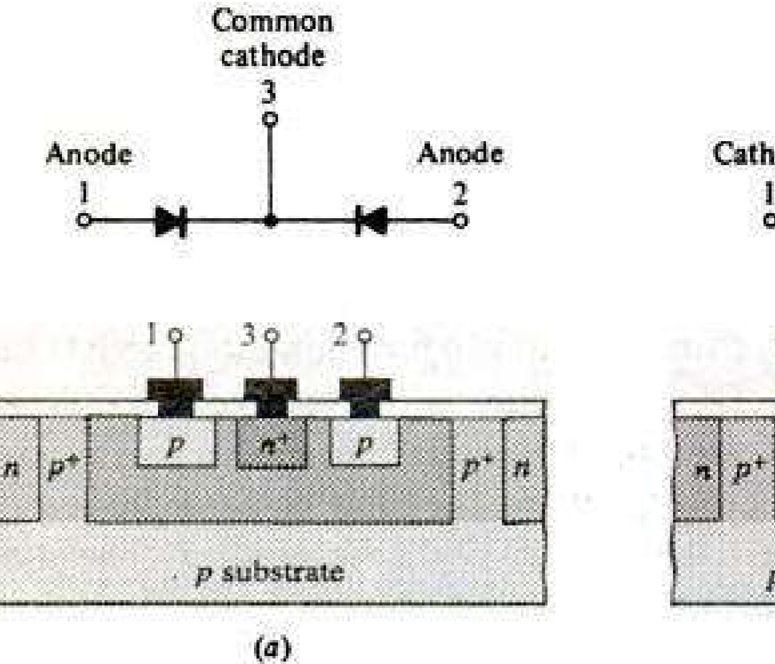



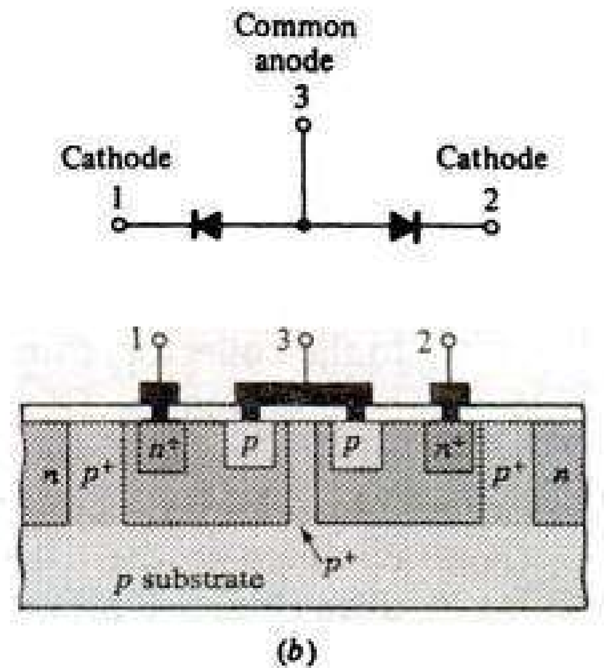

**FIGURE 5-17**
Cross section and connections for IC diodes: (*a*) emitter-base diode with collector short-circuited to base; (*b*) emitter-base diode with collector open; (*c*) collector-base diode (no emitter diffusion or implant).

(or not fabricated at all) (Fig. 5-17*c*). The choice of the diode type used depends on the application and circuit performance desired. Collector-base diodes have the higher collector-base voltage-breakdown rating of the collector junction (~12 V minimum), and they are suitable for common-cathode diode arrays diffused within a single isolation island, as shown in Fig. 5-18*a*. Common-anode arrays can also be made with the collector-base diffusion, as shown in Fig. 5-18*b*. A separate isolation is required for each diode, and the anodes are connected by metallization.

The emitter and base regions are very popular for the fabrication of diodes, provided the reverse-voltage requirement of the circuit does not exceed the lower base-emitter breakdown voltage (~7 V). Common-anode arrays can easily be made with the emitter and base diffusions by using the multiple-emitter

**FIGURE 5-18**
Diode pairs: (*a*) common-cathode, (*b*) common-anode.



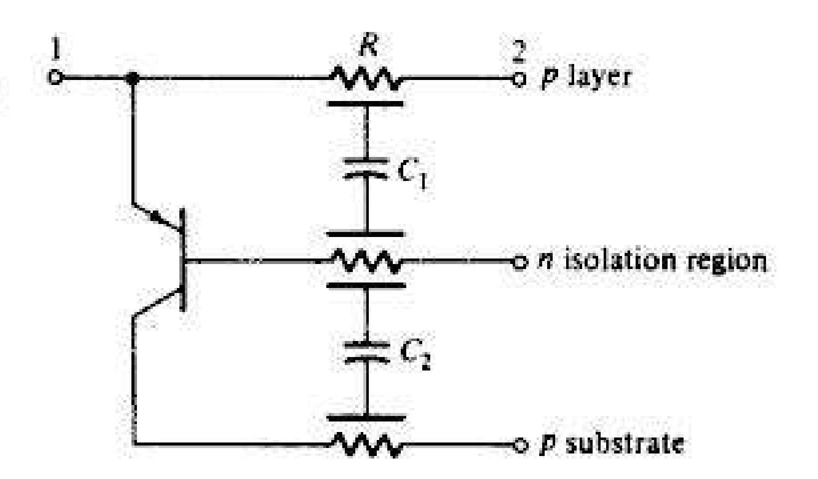



**FIGURE 5-23** The equivalent circuit of a diffused resistor.

of 50 μm, matching tolerance is about 0.2 percent. For this reason, IC design should, where possible, utilize *resistance ratios rather than absolute values*. The values of diffused resistors increase with temperature. For base-diffused resistors, this variation is in the order of 2000 ppm/°C (parts per million per degree Celsius); variations of 600 ppm/°C are typical of emitter-diffused resistors.

The equivalent circuit of the diffused resistor $R$ is shown in Fig. 5-23, where the parasitic capacitances of the base-isolation $C_1$ and isolation-substrate $C_2$ junctions are included. In addition, it can be seen that a parasitic *pnp* transistor exists, with the substrate as collector, the isolation $n$-type region as base, and the resistor $p$-type material as the emitter. The collector is reverse-biased because the $p$-type substrate is at the most negative potential. It is also necessary that the emitter be reverse-biased to keep the parasitic transistor at cutoff. This condition is maintained by placing all resistors in the same isolation region and *connecting the n-type isolation region surrounding the resistors to the most positive voltage present in the circuit*. Typical values of $\beta_F$ for this parasitic transistor range from 0.5 to 5.

**Ion-Implanted Resistors** As the base and emitter regions are often formed by ion implantation, this process is also used to form resistors having the same structure as shown in Fig. 5-22. Implanted $n$-type resistors can be fabricated by using an MOS process similar to that used to form the channel in an NMOS depletion transistor. Ion-implanted resistors are fabricated having resistance values comparable to those achieved with base diffusion. However, tolerances and temperature variations are well below those obtained for diffused resistors. Implanted resistance values can be controlled to three percent and temperature coefficients can be as low as 100 ppm/°C. Matching tolerances are also improved by about 25 percent compared with diffused resistors.

**Epitaxial Resistors** The sheet resistance of the collector epitaxial region is about six times that of the base diffusion. Hence it is possible to fabricate higher value resistances by using the epitaxial layer. Such a resistor is defined by the isolation diffusion which surrounds the resistor (Fig. 5-22). These sidewall effects become important, and, to maintain accuracy of resistance values, the isolation diffusion must be carefully controlled. The temperature variation of epitaxial resistors

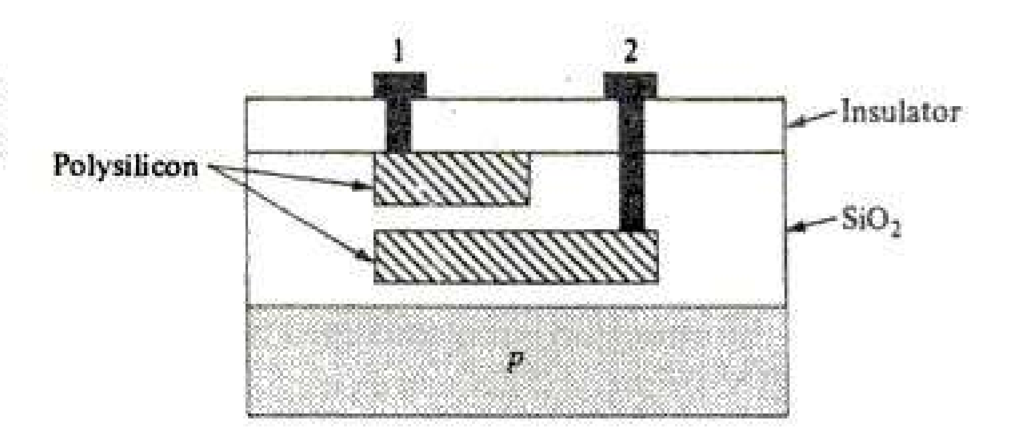



**FIGURE 5-27**
MOS capacitor formed from two polysilicon layers.

The capacitance of MOS or junction capacitors is quite small, generally about $4 \times 10^{-4}$ pF/$\mu$m$^2$. Thus a 40-pF capacitor requires an area of $10^5$ $\mu$m$^2$ or covers a rectangle $1 \times 0.1$ mm of chip surface. Most IC capacitors used are typically less than 100 pF. Values in excess of 500 pF have been obtained, but only at the expense of using a large portion of the chip area.

The use of tantalum films can increase the capacitance per unit area by a factor of 10. A controlled growth of tantalum pentoxide ($Ta_2O_5$) is used for the dielectric, and metallic tantalum is deposited for the top plate (since aluminum is soluble in $Ta_2O_5$). The increased capacitance is obtained at the expense of additional processing steps.

**5-10 INTEGRATED-CIRCUIT PACKAGING** The fabrication sequence is completed when all the processes required to realize all the components and their interconnections are performed. Each wafer is cut into chips (Fig. 5-3), so that the individual microelectronic systems are obtained. The chips are then placed into the package in which they will be used. Wire bonding connects the leads from the chip to the pins of the package. These are the terminals to which external components can be connected to the ICs. In general, external connections are determined by the system in which the circuits are used. Signal inputs and outputs, supply voltages, ground, and components that are not fabricated on the chip are typically applied to the external connections.

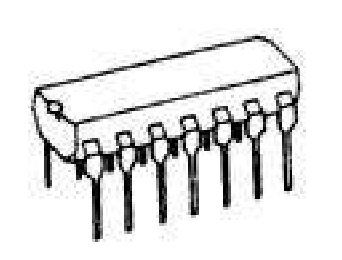


**FIGURE 5-28**
Dual-in-line (DIP) IC package.

A commonly used IC package is the *dual-in-line* (DIP) package shown in Fig. 5-28. The type of standard package can contain as few as 8 pins and as many as 40 pins. The particular number of pin connections is determined by the function the circuit must perform. In any given IC, however, not all pin connections are utilized.

**5-11 CHARACTERISTICS OF INTEGRATED-CIRCUIT COMPONENTS** From our discussion of IC technology, we can summarize the significant characteristics of ICs as follows:

1. Standard ICs (those stocked by manufacturers) are very inexpensive; for example, the National Semiconductor LM741 Op-Amp containing 21 transistors, 1 diode, and 12 resistors sells for less than 50 cents (in quantity).

5-8 How are the components interconnected in an IC?

5-9 Describe the ion-implantation process.

5-10 Sketch the cross section of an IC bipolar transistor.

5-11 Define buried layer. Why is it used?

5-12 Describe a lateral *pnp* transistor. Why is its current gain low?

5-13 Describe a vertical *pnp* transistor. Why is it of limited use?

5-14 Describe a super-$\beta$ transistor.

5-15 Sketch the cross section of an *n*-channel JFET.

5-16 Sketch the cross section of an enhancement-type NMOS transistor.

5-17 Repeat Rev. 5-16 for a depletion-type NMOS transistor.

5-18 (*a*) What is meant by polysilicon?
(*b*) What is the effect of using a polysilicon gate?

5-19 Sketch the cross section of a CMOS composite transistor.

5-20 (*a*) How are IC diodes fabricated?
(*b*) Draw the circuit diagram of two types of emitter-base diodes.

5-21 Sketch the top view of a multiple-emitter transistor. Show the isolation, collector, base, and emitter regions.

5-22 How is an aluminum contact made with *n*-type silicon so that it is (*a*) ohmic; (*b*) rectifying?

5-23 Why is storage time eliminated in a metal-semiconductor diode?

5-24 What is a Schottky transistor? Why is storage time eliminated in such a transistor? Are there any extra fabrication steps required to produce such a transistor? Explain.

5-25 Sketch the cross section of an IC Schottky transistor.

5-26 (*a*) Define sheet resistance $R_S$.
(*b*) Sketch the cross section of an IC resistor.
(*c*) What are the order of magnitudes of the smallest and the largest values of an IC resistance?

5-27 (*a*) Sketch the equivalent circuit of a base-diffused resistor, showing all parasitic elements.
(*b*) What must be done (externally) to minimize the effect of the parasites?

5-28 Describe a thin-film resistor.

5-29 (*a*) Sketch the cross section of a junction capacitor.
(*b*) Draw the equivalent circuit, showing all parasitic elements.

5-30 Repeat Rev. 5-29 for an MOS capacitor.

5-31 What are the two basic distinctions between a junction and an MOS capacitor?

5-32 (*a*) To what voltage is the substrate connected? Why?
(*b*) Repeat (*a*) for the isolation islands containing the resistors.
(*c*) Can several transistors be placed in the same isolation island? Explain.

5-33 List six important characteristics of integrated components.

5-34 List six design rules for monolithic-circuit layout.

6. Repeat steps 4 and 5 until the resultant quotient is zero. The digits in the *B* row, when read left to right, are the binary representation of the decimal number *D*. The leftmost bit, representing the place value of the highest power of 2, is the *most-significant bit* (MSB), whereas the rightmost bit is the *least-significant bit* (LSB).

*Example 6-1* Convert the decimal number 73 into binary.

*Solution* Set up an array similar to Table 6-1, as shown in Table 6-2. The binary representation of 73 is the 7-bit number 1001001. This result is readily checked as

$$1001001 = 1 \times 2^6 + 0 \times 2^5 + 0 \times 2^4 + 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 = 64 + 8 + 1 = 73$$

**TABLE 6-2 Array for Example 6-1**

| *8* | *7* | *6* | *5* | *4* | *3* | *2* | *1* | |
|---|---|---|---|---|---|---|---|---|
| 0 | $\frac{1}{2} = 0$ | $\frac{2}{2} = 1$ | $\frac{4}{2} = 2$ | $\frac{9}{2} = 4$ | $\frac{18}{2} = 9$ | $\frac{36}{2} = 18$ | $\frac{73}{2} = 36$ | $D = 73$ |
| 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | *B* |

The method outlined in Table 6-1 can be extended to convert a decimal number *D* to a base number *B*. The remainders $R_1, R_2, \ldots, R_n$, read from left to right, form the digits of the base *B* number. If, $B = 5$, for example, *R* can only have the value 0, 1, 2, 3, or 4. Often a subscript designating the base used to express a number is indicated. Thus $N_{10}$ is a decimal number and $N_2$ is a binary number.

Just as the decimal point separates positive and negative powers of 10, the binary point separates place values of positive and negative powers of 2. The binary number $101.011 = 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 0 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3}$ and has a decimal equivalent of 5.375.

Negative numbers are represented by adding a *sign bit* to the left of the MSB in a binary word. A zero (0) designates a positive number and a one (1), a negative number. Thus $0_\wedge1001001$ has the decimal equivalent +73, and $1_\wedge1001001$ denotes −73. The caret (∧) is sometimes used to indicate that the first bit is the sign bit.

A variety of binary-derived number representations and binary codes are used in digital systems. Several of these are introduced in Sec. 7-3 when we treat binary arithmetic.

## 6-2 BOOLEAN ALGEBRA

*Boolean algebra* is a two-state (binary) symbolic logic. A boolean variable *A* assumes one of two permissible values, 0 or 1. Thus *A* may be 1 ($A = 1$) or *A* may be 0 ($A = 0$). If *A* is not 1, *A* must be zero. A self-consistent

We shall have occasion to refer to the last two equations later in the Chapter.

The circuit in Fig. 2-13, analyzed in Example 2-2, is a diode-resistor implementation of a positive-logic AND gate. Reversing the diodes is all that is required to realize a negative-logic AND gate.

Note that the same circuit (Fig. 2-13) can be used to implement both the positive-logic AND gate and the negative-logic OR gate. This results from the fact that $V(0)$ in positive logic and $V(1)$ in negative logic each signify the lower of the two voltage levels. Similarly, $V(1)$ and $V(0)$ are the higher voltages in positive and negative logic, respectively. Thus we conclude that a *negative* OR *gate is the same circuit as a positive* AND *gate*. This result is not restricted to diode logic; it is valid independently of the hardware used to implement the circuit.

**The NOT (Inverter) Gate** The NOT circuit has a single input and a single output and performs the operation of *logic negation* in accordance with the following definition: *The output of a* NOT *circuit takes on the* 1 *state if and only if the input does* not *take on the* 1 *state*. The standard to indicate a *logic negation* is a small circle drawn at the point where a signal line joins a logic symbol. Negation at the input of a logic block is indicated in Fig. 6-5*a* and at the output in Fig. 6-5*b*. The symbol for a NOT gate and the boolean expression for negation are given in Fig. 6-5*c*. The equation is to be read "$Y$ equals NOT $A$" or "$Y$ is the complement of $A$." [Sometimes a prime (′) is used instead of the bar ( – ) to indicate the NOT operation.] The truth table is given in Fig. 6-5*d*.

A circuit which accomplishes a logic negation is called a NOT *circuit*, or, since it inverts the sense of the output with respect to the input, it is also known as an *inverter*. In a truly binary system only two levels $V(0)$ and $V(1)$ are recognized, and the output, as well as the input, of an inverter must operate between these two voltages. When the input is at $V(0)$, the output must be at $V(1)$, and vice versa. Ideally, then, a NOT circuit inverts a signal while preserving its shape and the binary levels between which the signal operates.

The bipolar and field-effect transistor (FET) switches discussed in Secs. 3-8 and 4-12 are inverters. We observed in Figs. 3-29 and 4-25 that an input transition from a low voltage to a high voltage produced the opposite transition in the output signal. Furthermore, the ideal controlled sources, introduced in Secs. 3-1 and 4-1, are inverter circuits when acting as controlled switches.

$A$ —○[ $Y = \bar{A}$ ]  (*a*)

[ $A$ ]○— $Y = \bar{A}$  (*b*)

$A$ —▷○— $Y$, $Y = \bar{A}$  (*c*)

| Input $A$ | Output $Y$ |
|---|---|
| 0 | 1 |
| 1 | 0 |

(*d*)

**FIGURE 6-5**
Logic negation at (*a*) the input and (*b*) the output of a logic block, (*c*) circuit symbol for an inverter (NOT gate), (*d*) the truth table.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

logic equation so as to transform it into a form which is better from the point of view of implementation in hardware. In the next section we shall verify through the use of boolean algebra that the four expressions given above for the exclusive-OR are equivalent.

**De Morgan's Laws** The following two binary equations are known as De Morgan's theorems:

$$\overline{ABC\cdots} = \bar{A} + \bar{B} + \bar{C} + \cdots \qquad (6\text{-}24)$$

$$\overline{A + B + C + \cdots} = \bar{A}\bar{B}\bar{C}\cdots \qquad (6\text{-}25)$$

To verify Eq. (6-24), note that if all inputs are 1, each side of the equation equals 0. On the other hand, if one (or more than one) input is 0, each side of Eq. (6-24) equals 1. Hence, for all possible inputs the right-hand side of the equation equals the left-hand side. Equation (6-25) is verified in a similar manner. De Morgan's laws complete the list of basic boolean identities. For future reference, all these relationships are summarized in Table 6-3.

With the aid of boolean algebra we shall now demonstrate the equivalence of the four exclusive-OR circuits on the preceding page. Using Eq. (6-24), it is immediately clear that Eq. (6-19) is equivalent to Eq. (6-23). Now the latter equation can be expanded with the aid of Table 6-3 as follows:

$$(A + B)(\bar{A} + \bar{B}) = A\bar{A} + B\bar{A} + A\bar{B} + B\bar{B} = B\bar{A} + A\bar{B} \qquad (6\text{-}26)$$

This result shows that the exclusive-OR in Eq. (6-20) is equilvalent to that in Eq. (6-23).

**TABLE 6-3 Summary of Basic Boolean Identities**

*Fundamental laws*

| OR | AND | NOT |
|---|---|---|
| $A + 0 = A$ | $A0 = 0$ | $A + \bar{A} = 1$ |
| $A + 1 = 1$ | $A1 = A$ | $A\bar{A} = 0$ |
| $A + A = A$ | $AA = A$ | $\bar{\bar{A}} = A$ |
| $A + \bar{A} = 1$ | $A\bar{A} = 0$ | |

*Associative laws*

$(A + B) + C = A + (B + C)$ $\quad (AB)C = A(BC)$

*Commutative laws*

$A + B = B + A$ $\quad AB = BA$

*Distributive law*

$A(B + C) = AB + AC$

*De Morgan's laws*

$\overline{AB\cdots} = \bar{A} + \bar{B} + \cdots$

$\overline{A + B + \cdots} = \bar{A}\bar{B}\cdots$

*Auxiliary identities*

$A + AB = A$ $\quad A + \bar{A}B = A + B$

$(A + B)(A + C) = A + BC$

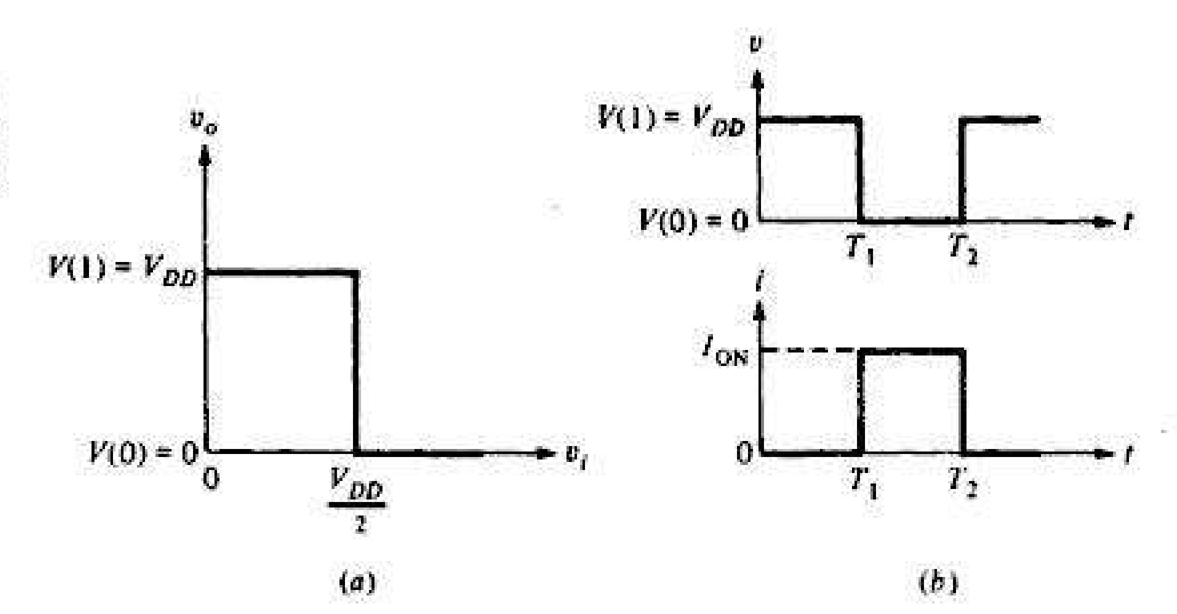



**FIGURE 6-14**
(*a*) Transfer characteristic and (*b*) voltage and current waveforms of an ideal inverter.

The output current and voltage waveforms corresponding to transitions from $V(1)$ to $V(0)$ and back to $V(1)$ that exist in an ideal controlled switch are shown in Fig. 6-14*b*. The static power dissipated in either state is zero because when $v_o \neq 0$, $i = 0$ (the ideal open switch) and when $i \neq 0$, $v_o = 0$ (the ideal closed switch). In addition, because the transition between states is instantaneous (Fig. 6-14*b*), the dynamic power dissipation, that is, the power consumed during switching, is also zero.

Two other characteristics of the ideal inverter are related to the interconnection of these gates: (1) the input circuit of the ideal inverter has no loading effect on the driving signal (the output of a previous gate), and (2) the inverter output is capable of driving an arbitrary number of similar gates with no degradation in output level.

The ideal inverter characteristics apply equally to multiple-input gates (AND, OR, NAND, NOR). In addition, such ideal gates can accept an arbitrary number of inputs without loading effects disturbing the driving stages. The preceding discussion leads us to observe that the following characteristics are important in evaluating the performance of a practical gate:

1. The range of voltages which correspond to the logic levels $V(0)$ and $V(1)$
2. The uncertainty region or range of input voltages for which the output state is undefined
3. The switching speed
4. Static and dynamic power dissipation
5. Input and output loading effects

**The Practical Inverter Transfer Characteristic** The transfer characteristic of a real, commercially available inverter circuit in Fig. 6-15 indicates several deviations from the ideal characteristic in Fig. 6-14*a*. These differences are that the voltages $V(1)$ and $V(0)$ are not constant, that they can differ from the supply voltage and zero, respectively, and that the transition between states is not abrupt. The general shape of the transfer curve is similar to the transfer characteristics

 6-4 Logic Gate Characteristics 

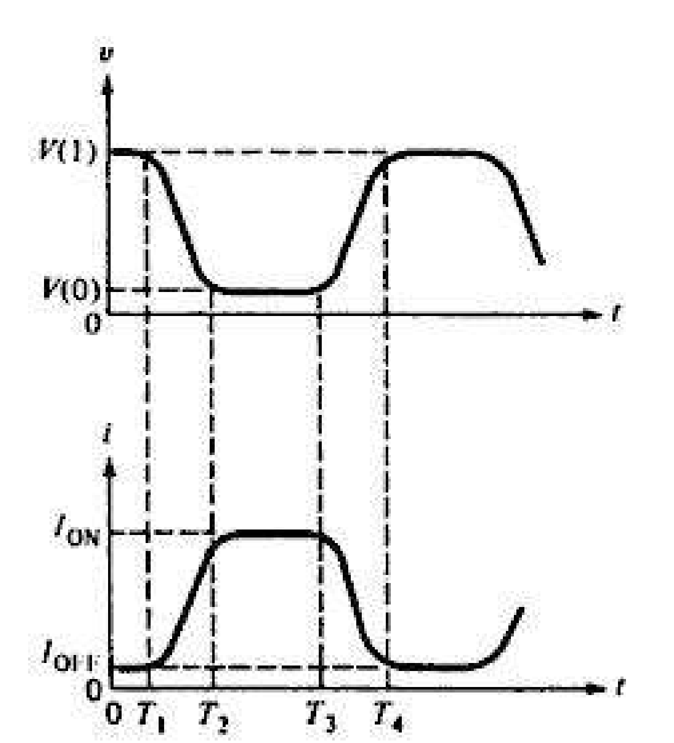



**FIGURE 6-18**
Practical current and voltage waveforms in a switch. Power dissipation exists in both logic states (static) and during the transition between states (dynamic).

Both static and dynamic dissipation contribute to the total power consumption of a gate. Sometimes static dissipation is the dominant factor. In other instances such as certain VLSI systems fabricated in CMOS technology, dynamic power dissipation is the major portion of total power consumption.

**Speed of Operation** The speed at which a gate can be operated is influenced by the time for a signal to propagate from input to output and the transition time between states. In Fig. 6-19 we display typical input and output waveforms of an inverter. The *rise* and *fall* times $t_r$ and $t_f$, respectively, are measures of the transition times between logic states. Both terms are defined by the time difference between

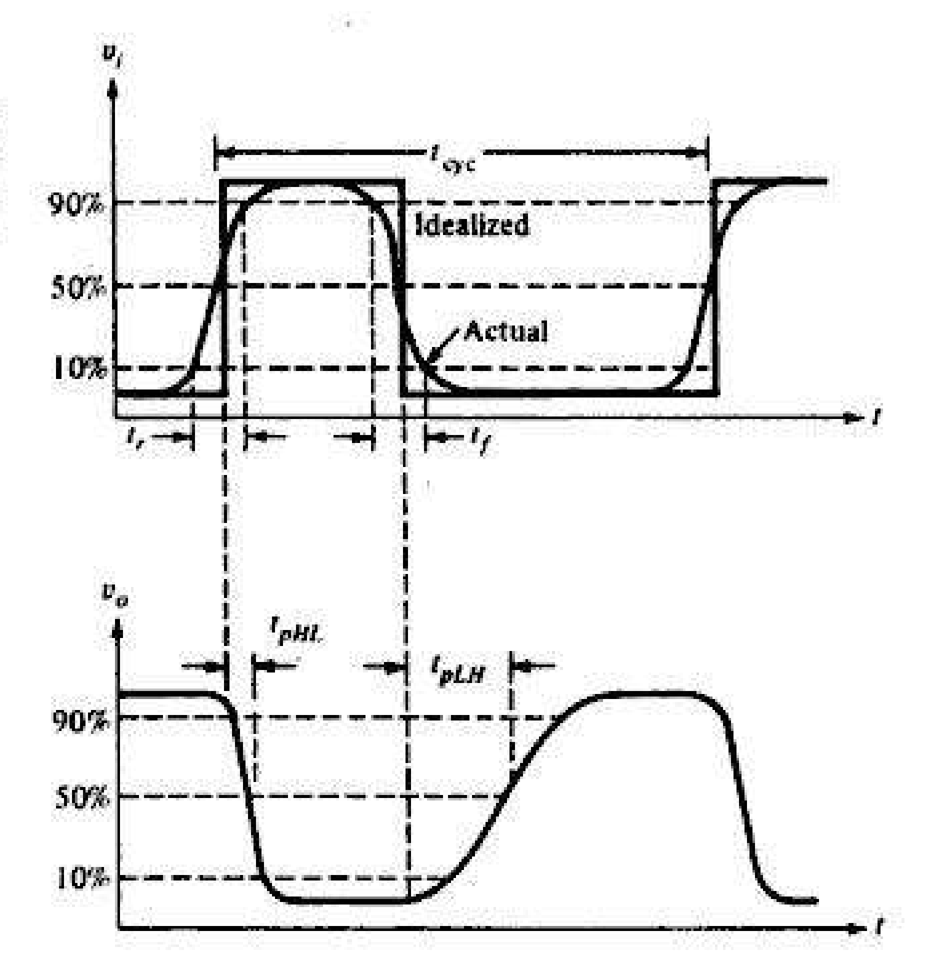



**FIGURE 6-19**
Input and output waveforms for one cycle showing rise time, fall time, cycle time, and propagation delay.

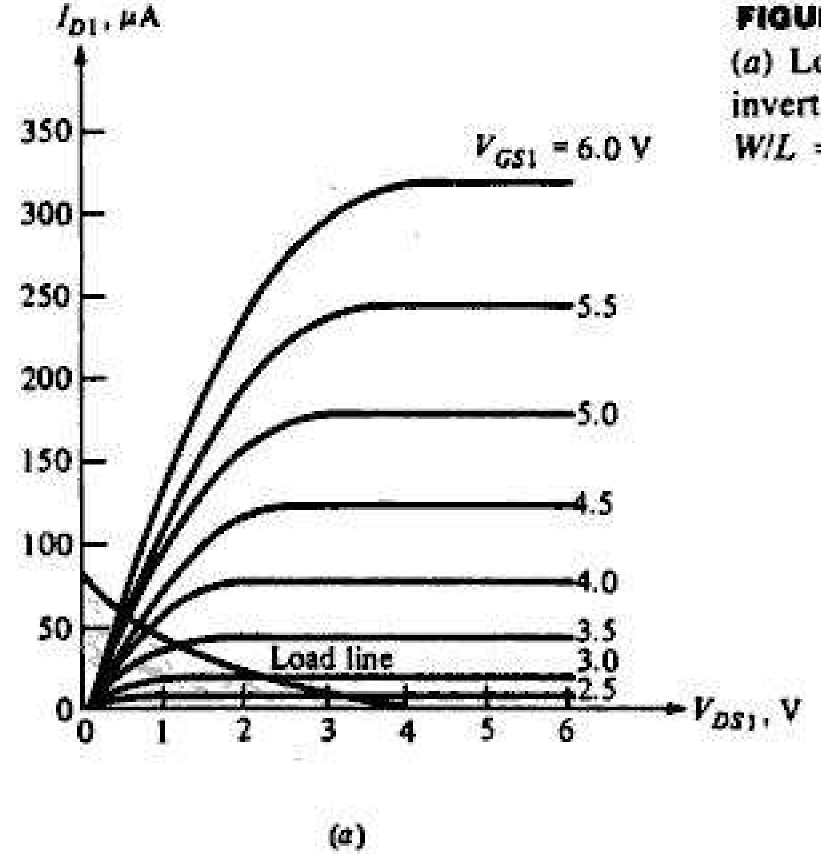
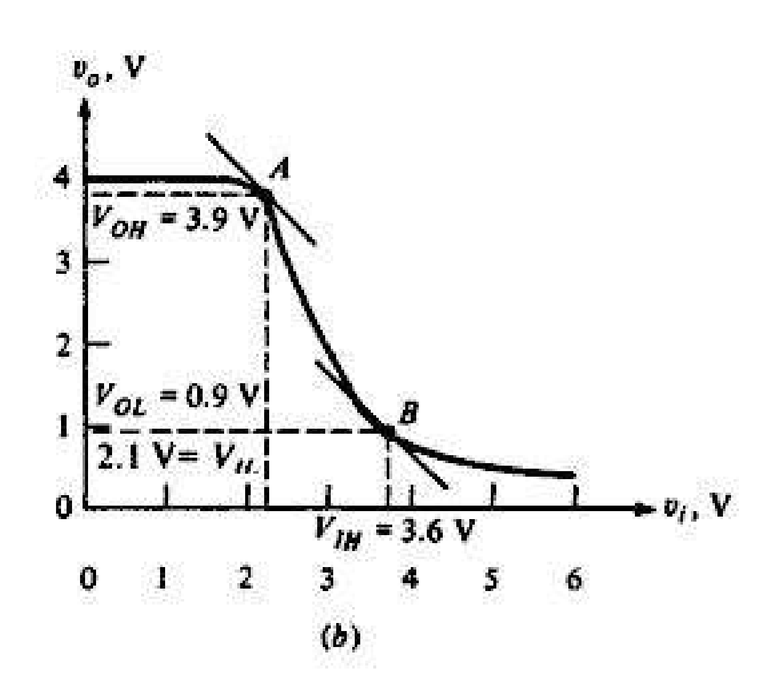



**FIGURE 6-22**
(*a*) Load line and (*b*) voltage transfer characteristic for inverter in Example 6-4. The enhancement load has $W/L = \frac{1}{4}$ and the aspect ratio of the driver is unity.




characteristics in the ohmic region than does that in Fig. 6-21*a* and causes the transition between states to be more abrupt.

Further decrease in the aspect ratio results in improved noise margins and more sharply delineated logic levels. However, this can be achieved only at the expense of additional chip area as the gate length of $Q2$ must be increased. Circuit designers, therefore, must make the trade-off between improved performance and reduced component densities. Current practice indicates that aspect ratios of $\frac{1}{4}$ or $\frac{1}{5}$ provide acceptable design compromises.

**The Linear (Nonsaturated) Load** A second technique for improving the performance of an NMOS inverter is to use a linear (ohmic or nonsaturating) load resistance. From the discussion in Sec. 4-7 it follows that a MOSFET will operate in the ohmic region if

$$V_{GS} - V_{DS} > V_T \tag{6-30}$$

In Fig. 6-23*a*, the gate of $Q2$ is connected to a separate supply $V_{GG} = 9$ V. Applying KVL around the loop $G_2 - S_2 - D_2 -$ ground $- G_2$, we obtain

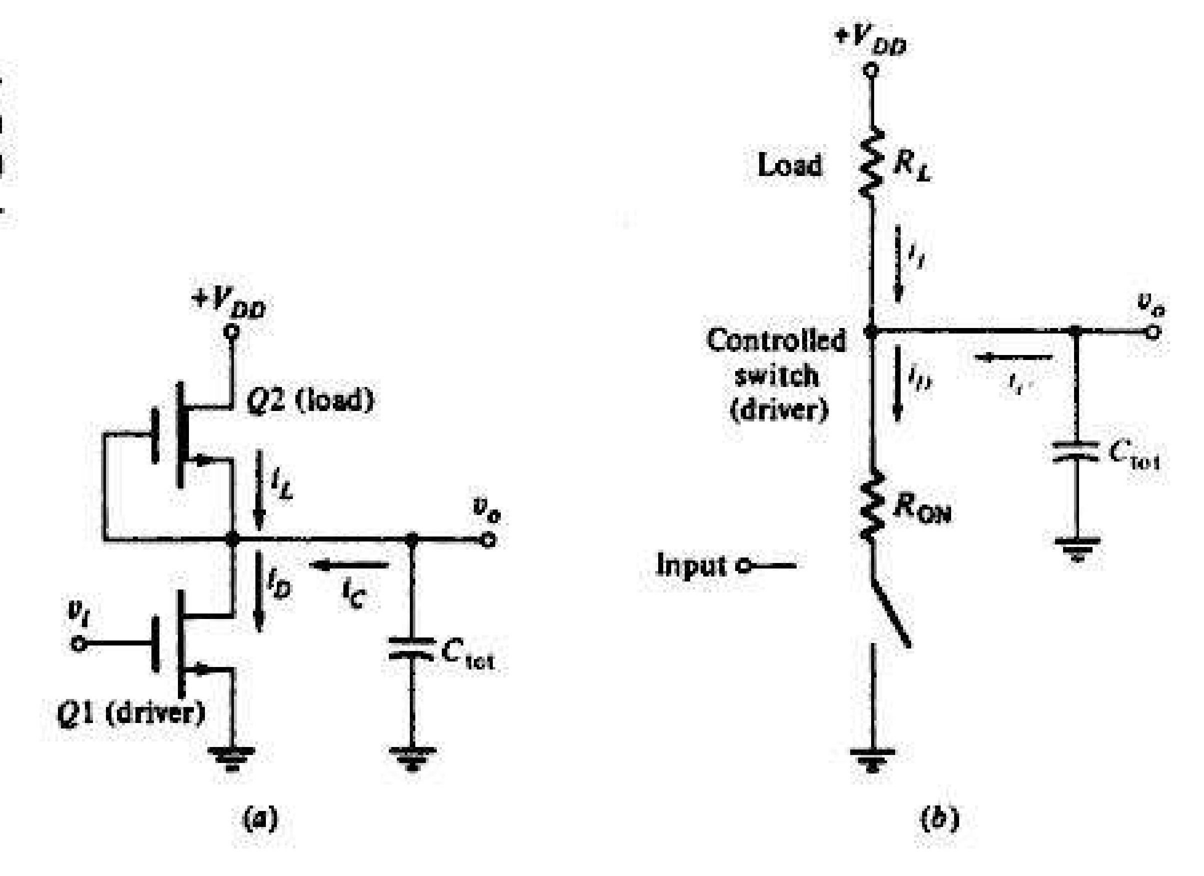
$$V_{GS2} - V_{DS2} + V_{DD} - V_{GG} = 0 \tag{6-31}$$

Hence $V_{GS2} - V_{DS2} = V_{GG} - V_{DD} = 9 - 6 = 3$ V, and, for $V_T = 2$ V, Eq. (6-30) is satisfied so that $Q2$ is constrained to operate in its linear region.

Consider the inverter shown in Fig. 6-23, in which $Q1$ has the characteristics illustrated in Fig. 6-21*a* and $Q2$ has the characteristics displayed in Fig. 6-23*b*. From Eq. (6-31) we find

$$V_{DS2} = V_{GS2} - 3 \tag{6-32}$$

**FIGURE 6-26**
(*a*) An NMOS depletion-load inverter with capacitive load. (*b*) Equivalent circuit representation.




$V(1) = V_{DD}$; thus $C_{tot}$ is charged to $V_{DD}$. The switch is now closed at $t = 0$ because of an instantaneous input transition from $V(0)$ to $V(1)$. The output is required to make the opposite transition and $C_{tot}$ must be discharged toward $V(0)$. The equivalent circuit for $t \geq 0$ is displayed in Fig. 6-26*b* with the switch closed ($R_{ON}$ connected to ground). The capacitor will discharge to $V(0) = R_{ON} V_{DD}/(R_{ON} + R_L)$ with a time constant $\tau_{HL} = C_{tot} R_{ON} R_L/(R_{ON} + R_L)$.[1] The output waveform for $v_o(t)$ is shown in Fig. 6-27*a*. The propagation delay $t_{pHL}$ is defined, as indicated in Fig. 6-27*a*, as the time for $v_o$ to fall from $V(1)$ to the 50 percent voltage $V'$ [halfway between $V(1)$ and $V(0)$]. Note that

$$V' = V(0) + \tfrac{1}{2}[V(1) - V(0)] = \tfrac{1}{2}[V(1) + V(0)] \qquad (6\text{-}33)$$

[1]To verify these results, the reader need only obtain a Thévenin equivalent of the portion of the circuit containing $R_{ON}$, $R_L$, and $V_{DD}$.

**FIGURE 6-27**
Output waveforms for Fig. 6-26*b* showing (*a*) transition from $V(1)$ to $V(0)$ and (*b*) the transition from $V(0)$ to $V(1)$.

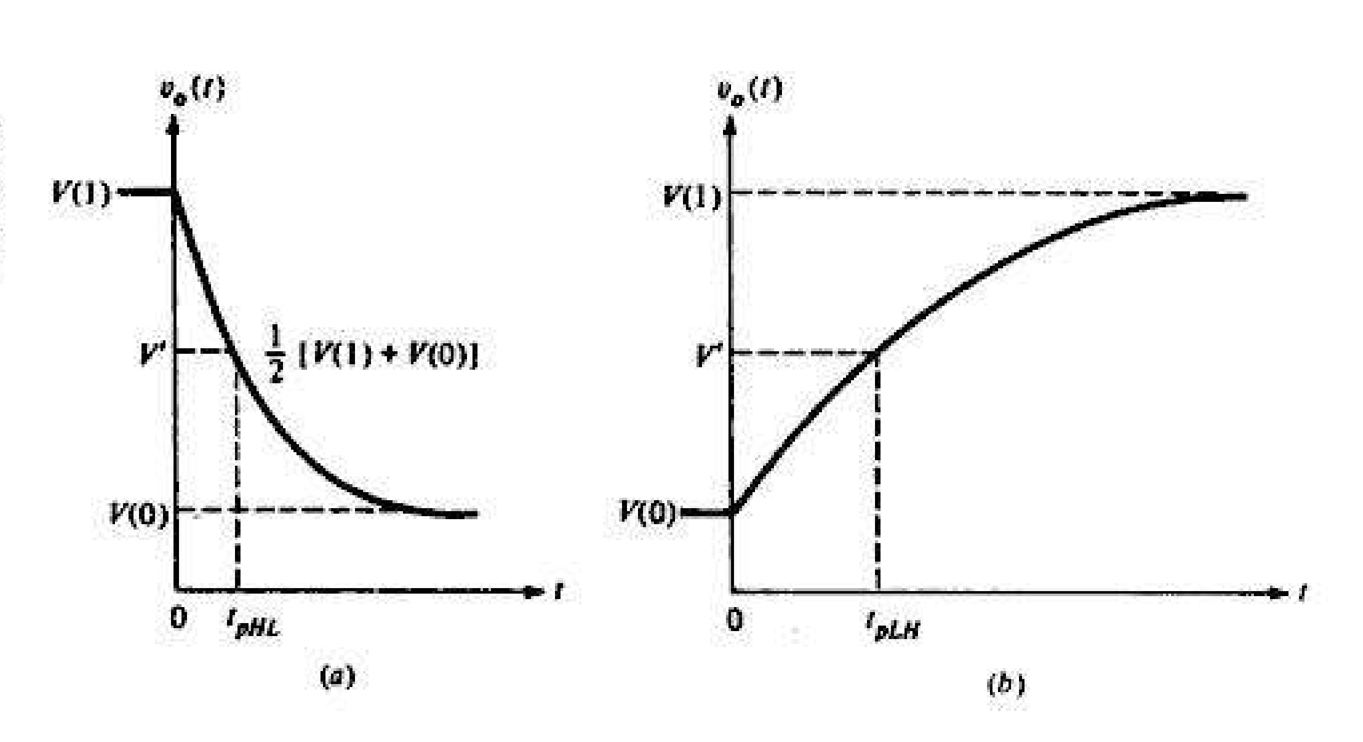
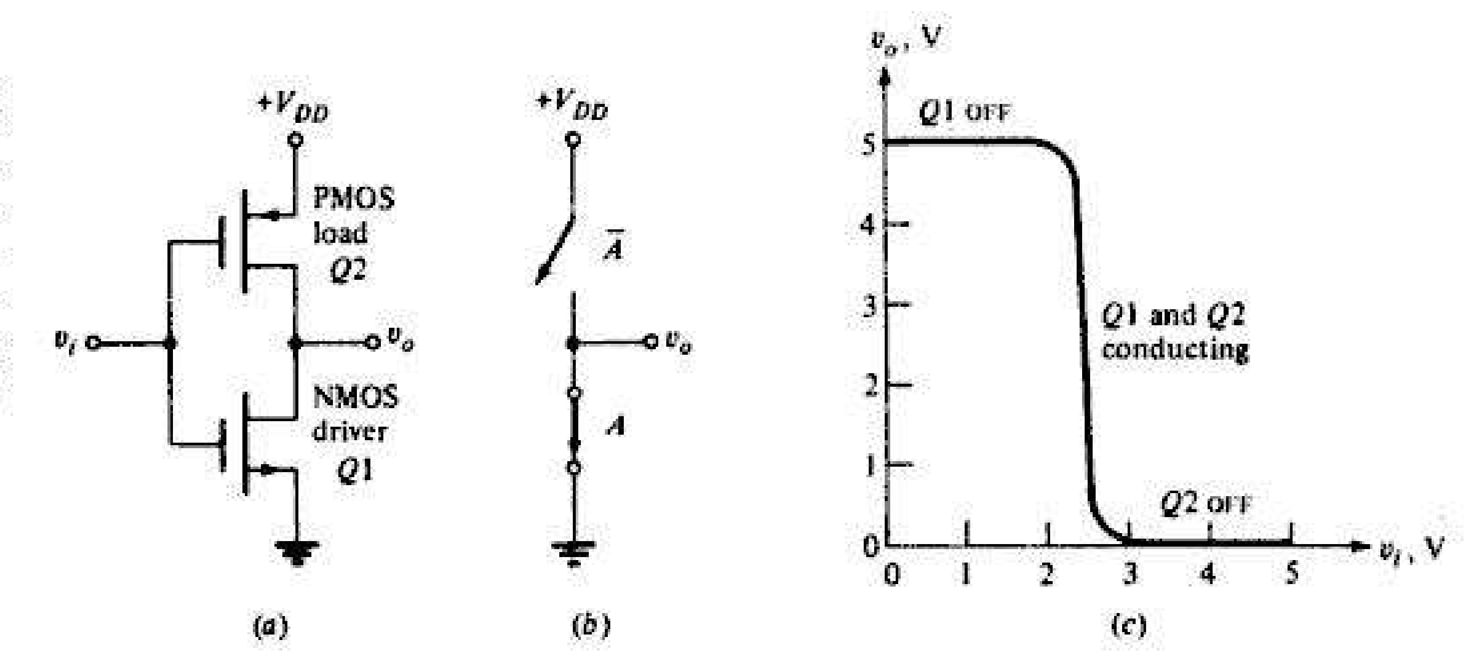


**FIGURE 6-30**
(*a*) The circuit diagram of a CMOS inverter and (*b*) its equivalent switch representation. (*c*) The voltage transfer characteristic for (*a*) with $V_{DD} = 5$ V and threshold voltages of 2 V ($Q1$) and $-2$ V ($Q2$).




**6-8 THE CMOS INVERTER** Complementary metal-oxide-semiconductor digital circuits are widely used because they have the distinct advantage of having virtually no static power dissipation in either the logic 1 or logic 0 state. The basic CMOS inverter, initially described in Sec. 4-15, has the circuit configuration shown in Fig. 6-30. The series-connected NMOS driver and PMOS load are both enhancement transistors. Their drains are connected, and the output signal is taken at this node. The input signal is applied simultaneously to both devices at the common gate terminal formed by connecting both gates. The input voltage $v_i$ varies from $V(0) = 0$ V to $V(1) = V_{DD}$. When $v_i = 0$, then $V_{GS1} = 0$ and $Q1$ is OFF, whereas $V_{GS2} = -V_{DD}$ and the PMOS device $Q2$ is ON. However, since the two FETs are in series, the current in $Q2$ equals that in $Q1$ ($I_{D1} = -I_{D2} = 0$), even though the gate voltage has a magnitude that nominally causes conduction. In other words, $Q2$ operates at the origin of the PMOS output characteristic corresponding to the gate voltage $V_{GS2} = -V_{DD}$. Since $V_{DS2} = 0$, it follows that $v_o = V_{DD}$. Inverter action has been verified because $v_o = V(1)$ for $v_i = V(0)$.

Consider now that $v_i = V_{DD} = V_{GS1}$; then $Q1$ is ON but $Q2$ with $V_{GS2} = 0$ is OFF. Hence $I_{D1} = -I_{D2} = 0$ and $Q1$ operates at the origin of the NMOS drain characteristics independent of $V_{GS1}$. Since the voltage across $Q1$ is zero, $v_o = 0$. Again the NOT property is obtained; $v_o = V(0)$ for $v_i = V(1)$. In either logic state $Q1$ or $Q2$ is OFF and the quiescent power dissipation is theoretically zero. In reality, the *standby power* equals the product of the OFF leakage current and $V_{DD}$ and equals a few nanowatts per gate.

From the discussion in the previous paragraphs, we can see that the switching circuit in Fig. 6-30*b* is analogous to CMOS operation. Because one switch must always be open, no current path exists between the supply voltage and ground. Consequently, the power is always zero. In the ideal situation, switching is instantaneous and no *dynamic power* is consumed. In the next paragraph we demonstrate, that the dynamic power is not zero in practice.

Consider the circuit in Fig. 6-30*a* for which $V_{DD} = 5$ V, $Q1$ has $V_T = 2$ V, and $Q2$ has $V_T = -2$ V. Let us assume that the processing factor $k$ and the

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

can also drive a load which is not on the same chip as the NOR gate (called an "off-chip" load) and has increased wiring capacitance.

The newest family of CMOS circuits is the 74HC series which utilizes more advanced fabrication methods to reduce device dimensions. Consequently, the propagation delays are reduced in comparison to other CMOS families. This higher-speed CMOS family is also capable of driving the input of a TTL 74LS gate and has comparable propagation delay. However, as with all CMOS circuits, static gate dissipation is extremely small. As the transfer characteristic, particularly with double-buffering, is most nearly that of the ideal logic circuit, noise margins are large and CMOS gates display excellent noise immunity.

## 6-10 THE BJT INVERTER

The BJT inverter, depicted in Fig. 6-34*a*, is simply the transistor switch described in Secs. 3-5, 3-6, and 3-8. The transfer characteristic of this circuit can be developed as follows:

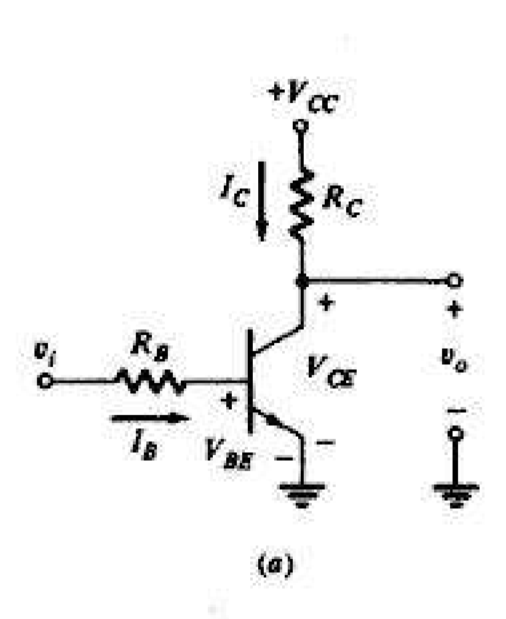
1. For $v_i \leq 0.5$ V $= V_\gamma$, the cut-in voltage (see Table 3-1), the BJT is cut off and, neglecting $I_{CO}$, the output voltage is $V_{CC}$.
2. The transistor is just barely conducting for $v_i$ slightly greater than 0.5 V, and the small collector current produced causes $v_o$ to decrease from $V_{CC}$ by $I_C R_C$.
3. In the forward-active region, $V_{BE(ON)} = 0.7$ V, $I_C = \beta_F I_B$, and

$$I_B = \frac{v_i - V_{BE(ON)}}{R_B} \tag{6-37}$$

The output voltage is given by

$$v_o = V_{CC} - I_C R_C = V_{CC} - \frac{\beta_F R_C}{R_B}[v_i - V_{BE(ON)}] \tag{6-38}$$

Consequently, from Eq. (6-38), $v_o$ decreases linearly with $v_i$.



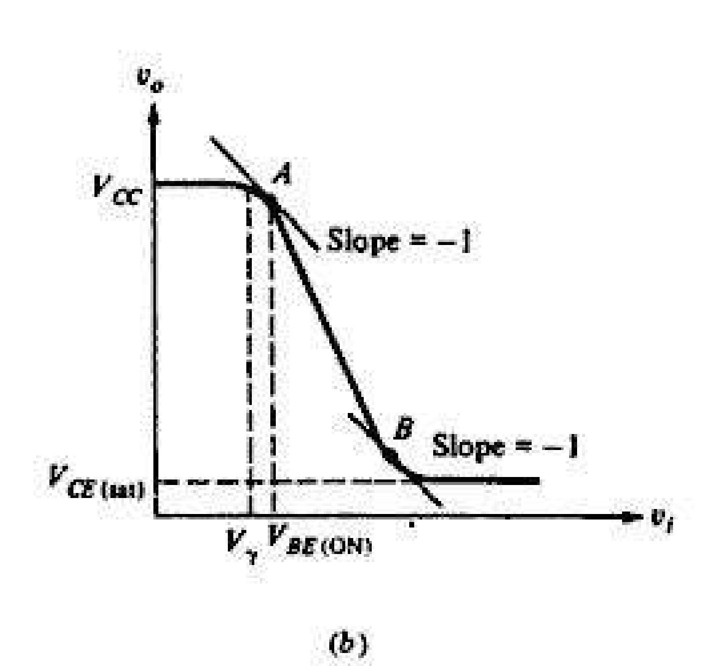




**FIGURE 6-34**
(*a*) The circuit and (*b*) the transfer characteristic for a BJT inverter.

The basic TTL NAND circuit is displayed in Fig. 6-37 and employs the topology of the DTL gate. The emitter junctions of the multiple-emitter transistor $Q1$ in Fig. 6-37 replaces the diodes $D$ in the left-hand rectangle in Fig. 6-36. Also, $D1$ is replaced by the collector junction of $Q1$. The emitter junction and emitter resistance $R_3$ of $Q2$ in Fig. 6-37 replace $D2$ and $R_2$ in Fig. 6-36. Both circuits utilize an output inverter ($Q3$ or $Q$).

The explanation of the operation of the TTL gate parallels that of the DTL switch. Thus, if at least one input is at $V(0) = 0.2$ V, then

$$V_P = 0.2 + 0.2 = 0.4 \text{ V}$$

For the collector junction of $Q1$ to be forward-biased and for $Q2$ and $Q3$ to be ON requires $V_P$ to be about $0.7 + 0.7 = 1.4$ V. Hence $Q2$ and $Q3$ are OFF; the output rises to $V_{CC} = 5$ V, and $Y = V(1)$. On the other hand, if all inputs are high (at 5 V), the input diodes (the emitter junctions) are reverse-biased and $V_P$ rises toward $V_{CC}$ and drives $Q2$ and $Q3$ into saturation. Then the output is $V_{CE(sat)} = 0.2$ V, and $Y = V(0)$ (and $V_P$ is clamped at about 1.6 V).

**Input Transistor Action** The explanation given in the preceding paragraph assumes that $Q1$ acts like isolated back-to-back diodes and not as a transistor. The preceding conclusions are also reached if the transistor behavior of $Q1$ is taken into consideration.

**Condition 1.** *At least one input is low,* $v_i = 0.2$ V. The emitter of $Q1$ is forward-biased and we assume that $Q2$ and $Q3$ are OFF. The current $I_{C1}$ ($= I$) into the collector $P$ must be the current from emitter to base of $Q2$. Hence, $I_{C1}$ equals the reverse saturation current of the emitter-junction diode of $Q2$. Since this current is very small (a few nanoamperes), $I_{B1} \gg I_{C1}/\beta_F$ and $Q1$ is in saturation. The voltage at $P$ equals $V_{CE(sat)} + v_i = 0.2 + 0.2 = 0.4$ V. This voltage is too small to put $Q2$ and $Q3$ ON. This argument justifies our assumptions that $Q2$ and $Q3$ are OFF and, therefore, $Y = V(1) = V_{CC}$.

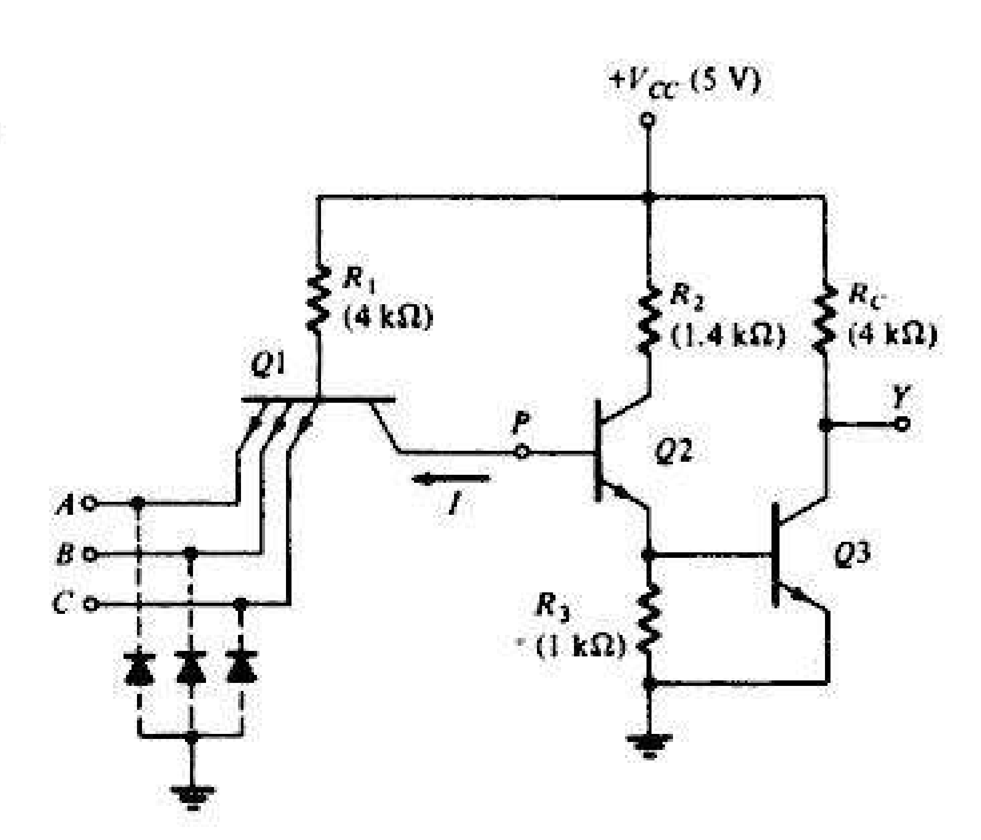



**FIGURE 6-37**
The basic TTL NAND-gate configuration.

*wire* OR-*ing* or as the OR-*tied connection*) and by enabling only the *i*th chip during the interval when $Y_{ni}$ is to appear at $Y_n$. The TTL totem-pole output stage (Fig. 6-38), modified to include such an enable, is indicated in Fig. 6-39*a* and the corresponding open-collector output circuit is shown in Fig. 6-39*b*.

In Fig. 6-39*a* if the *enable* or *chip-select* (*CS*) signal is low, *D*1 and *D*2 are OFF, and the output is either in state 1 or state 0, depending on whether the input data are 0 or 1. However, if *CS* is high, then *D*1 and *D*2 are ON, these diodes clamp *Q*3 and *Q*4 OFF, and the output *Y* is effectively an open circuit. This condition, referred to as the *high-impedance third state*, allows OR-ing of the outputs from the several packages. The circuit in Fig. 6-39*b* operates in a similar 3-state fashion. However, the manufacturers designate the configuration in Fig. 6-39*a* as the tristate (TS) output and that in Fig. 6-46*b* as the open-collector (OC) output.

**The Transfer Characteristic** The TTL NAND gate (Fig. 6-38) becomes an inverter when all the inputs are tied together. The piecewise linear approximation to the transfer characteristic of the resultant inverter, shown in Fig. 6-40, differs from that shown in Fig. 6-35 for the basic BJT inverter. The following qualitative argument justifies the shape of the characteristic in Fig. 6-40; the numerical evaluation of the critical voltage values is left to the student (Prob. 6-66).

For $v_i < V_{IL}$, both *Q*2 and *Q*3 are cut off, *Q*4 is in saturation, and the output is *V*(1). At point *A*, *Q*2 begins to conduct. However, the current produced in *Q*2 is insufficient to produce the voltage drop $V_{EN2} = V_{BE3}$ needed to turn on *Q*3. The accompanying decrease in $V_{CN2}$ maintains *Q*4 in a conducting state, but it is no longer in saturation and accounts for the decrease in $v_o$. Increasing $v_i$ to its value at *B* increases the emitter current in *Q*2 and hence turns on *Q*3. Between *B* and *C* in Fig. 6-40, *Q*3 is in its forward-active region and the output decreases with increasing $v_i$ (similar to the region between *A* and *B* in Fig.

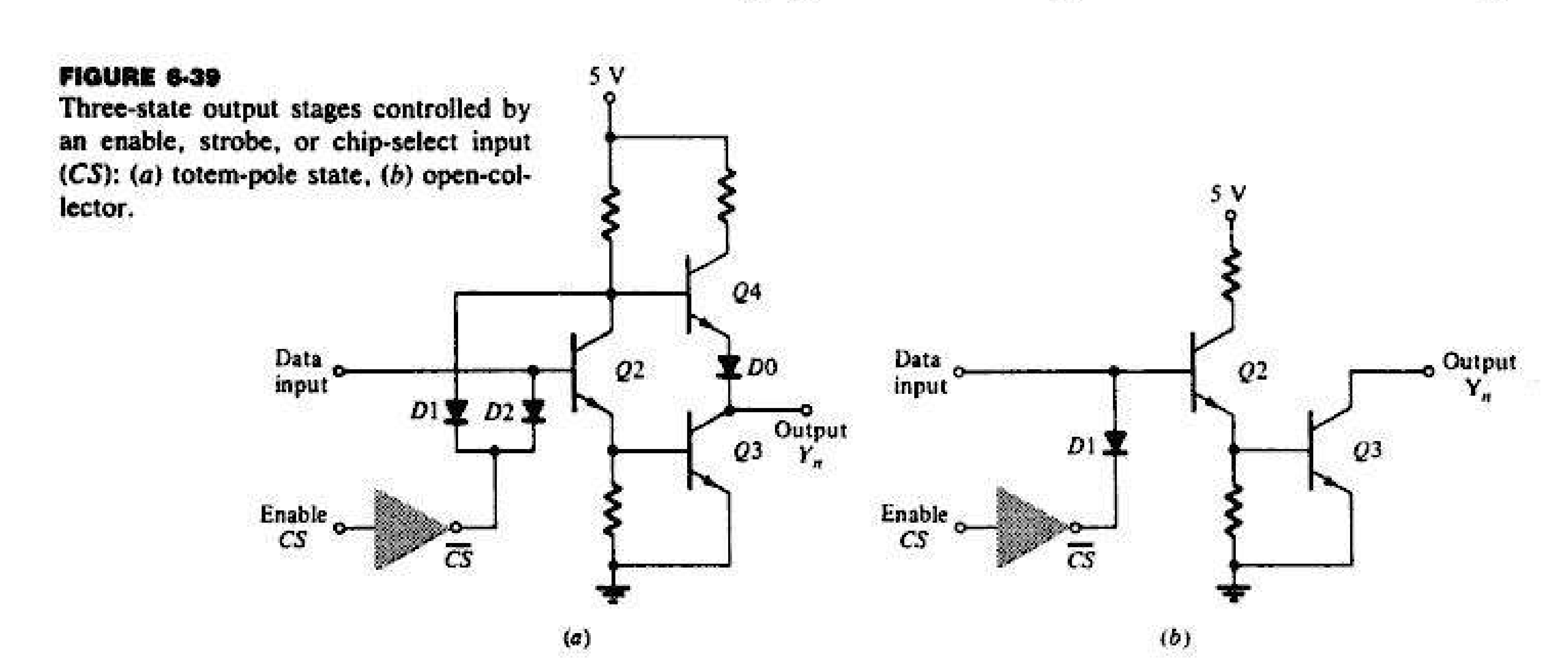



**FIGURE 6-39**
Three-state output stages controlled by an enable, strobe, or chip-select input (*CS*): (*a*) totem-pole state, (*b*) open-collector.

transistors $Q1$ and $Q2$, when conducting $(-0.1 < v_d < 0.1)$, are in the forward-active region and either $Q1$ or $Q2$ is virtually cut off when $v_d$ is outside this range.

Note that for $|v_d| \geq 4V_T$, either $v_{o1}$ is high and $v_{o2}$ is low or vice versa. In boolean variable terms, the two outputs are complementary (if $v_{o1} = Y$, then $v_{o2} = \overline{Y}$). Emitter-coupled logic gates exploit this convenience of providing both outputs $Y$ and $\overline{Y}$ simultaneously.

**The Basic ECL OR/NOR Gate** The standard ECL OR/NOR gate topology is displayed in Fig. 6-43*a*. It is obtained from Fig. 6-42*a* by using a constant reference voltage $V_R$ for $v_2$ of $Q2$ and by paralleling transistors which share a common collector resistance in place of $Q1$. A fan-in of 2 is depicted in Fig. 6-43*a*. If either input $A$ or $B$ is $V(1) > V_R + 0.1$, the outputs are $v_{o1} = V(0)$ and $v_{o2} = V(1)$. If both inputs are greater than $V_R + 0.1$, again $v_{o1} = V(0)$ and $v_{o2} = V(1)$. However, when both $A$ and $B$ are at $V(0) < V_R - 0.1$, $v_{o1} = V(1)$ and $v_{o2} = V(0)$. Consequently, $v_{o2}$ is the OR output ($Y_2 = A + B \equiv Y$) and $v_{o1}$ is the NOR output $[Y_1 = \overline{(A + B)} = \overline{Y}]$, as depicted in Fig. 6-43*b*.

**The Reference Voltage $V_R$** The current source in Fig. 6-43 is realized in its simplest form by a resistor $R_E$ placed between the common emitters and the negative supply voltage. In Fig. 6-43 (with $R_E$ replacing $I_{EE}$), KVL applied to the loop consisting of $V_R$, $V_{BE2}$, $R_E$, and $V_{EE}$ allows the current $I$ in $R_E$ to be expressed as $I = (V_R - V_{BE2} + V_{EE})/R_E$. Since the variation in $V_{BE2}$ with current in the active region is small compared with $V_R + V_{EE}$, $I$ remains essentially constant in the emitter-coupled pair. This constant current simulates the current source $I_{EE}$ in Fig. 6-42*a*.

The circuit in Fig. 6-43*a* requires three power supplies ($V_{CC}$, $V_{EE}$, and $V_R$). A more practical single-supply ($V_{EE}$) ECL gate is obtained by setting $V_{CC}$ to

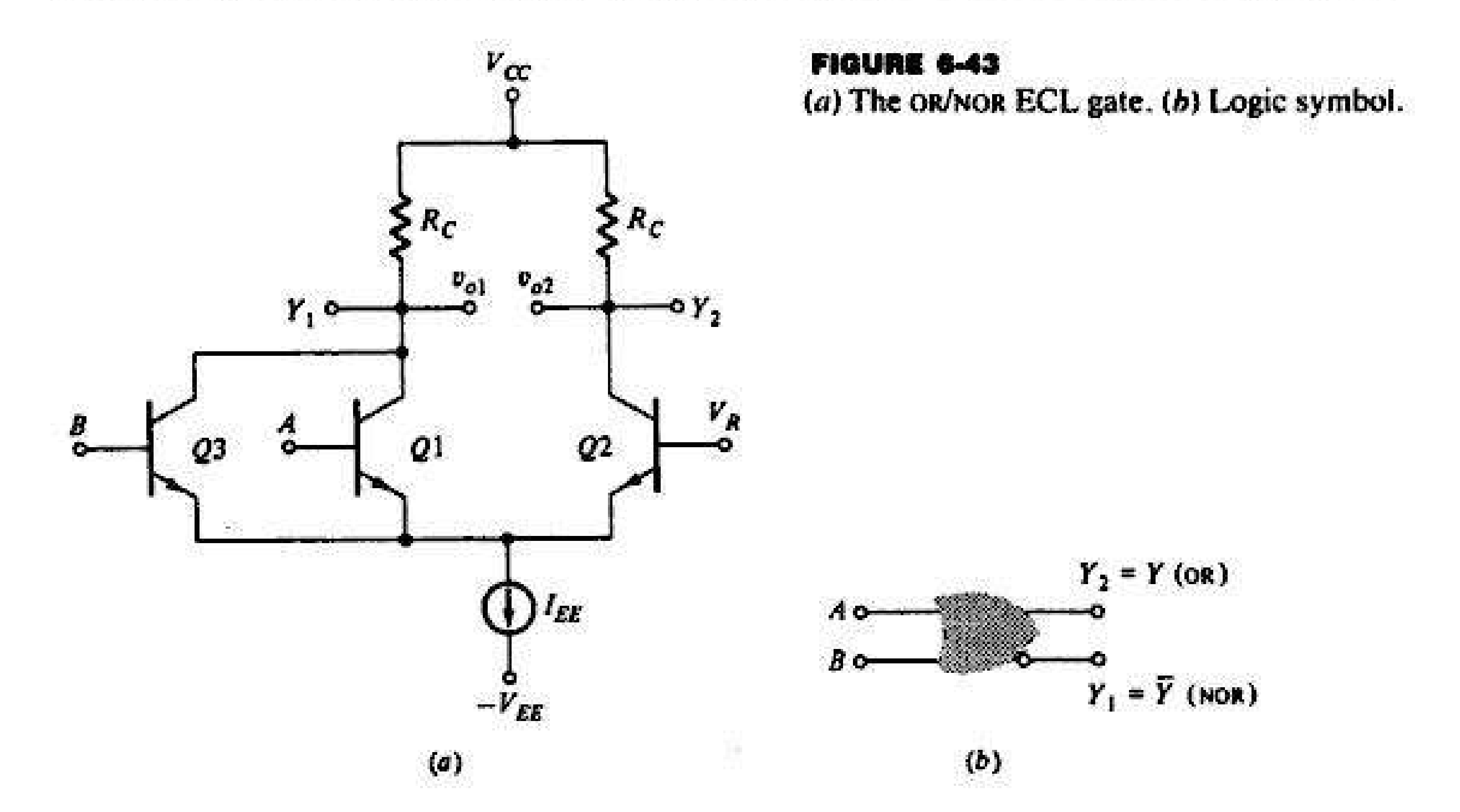



**FIGURE 6-43**
(*a*) The OR/NOR ECL gate. (*b*) Logic symbol.

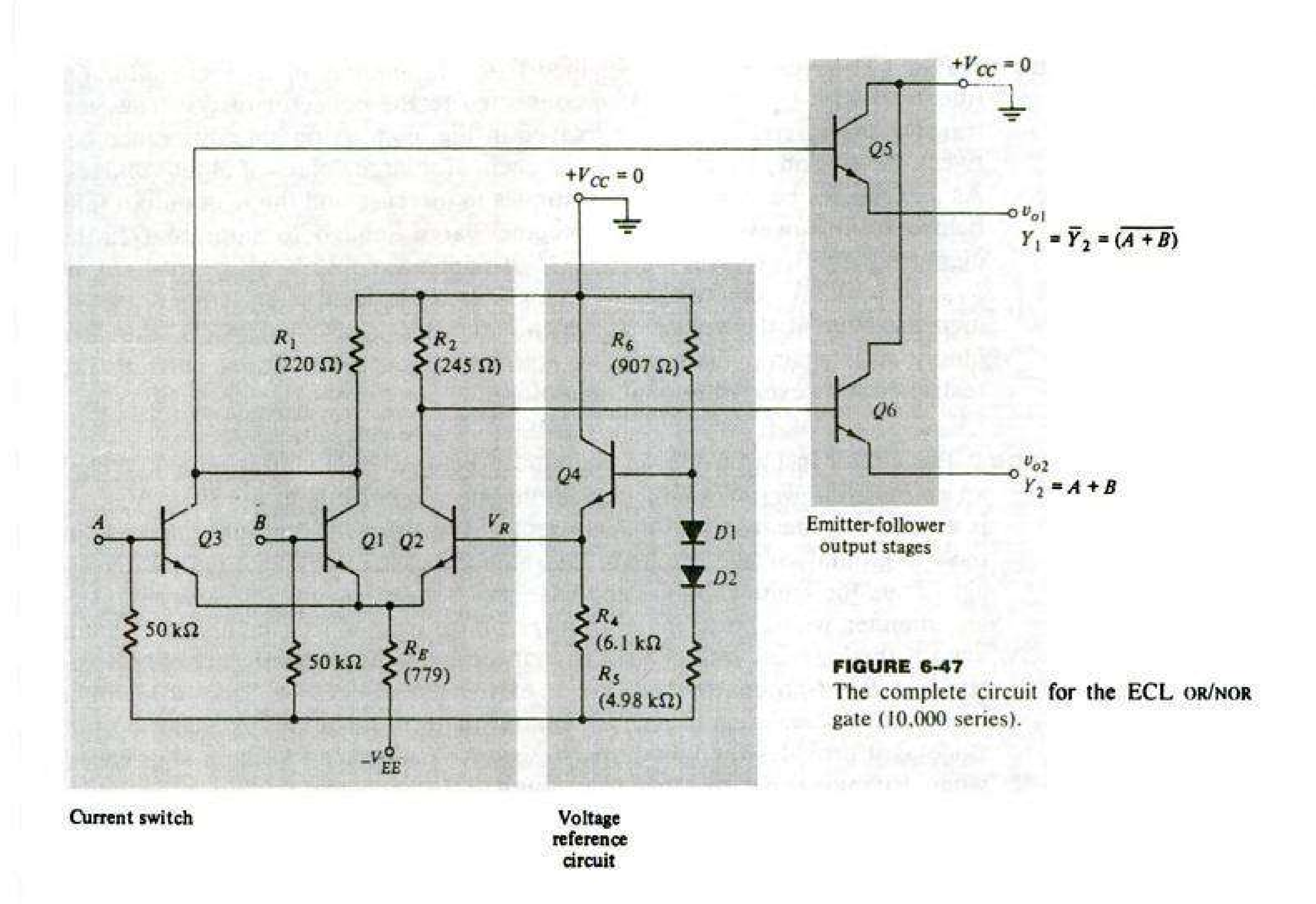



**FIGURE 6-47**
The complete circuit for the ECL OR/NOR gate (10,000 series).

**Noise Margins**

Determinations of $V_{IL}$ and $V_{IH}$ are usually made to correspond to differential input voltages that result in collector currents in the OFF transistor being 1 percent of the current in the ON transistor. The values of $V_{IL}$ and $V_{IH}$ so obtained differ only slightly from the values obtained at the unity-slope points (Prob. 6-74). With reference to Eq. (3-41), the change in $v_d$ necessary to make $I_{C1}/I_{C2} = 100$ is evaluated as $v_d \simeq 112$ mV. By symmetry, $V_{IH} = V_R + 0.112$ and $V_{IL} = V_R - 0.112$. Using $V_R = -1.32$ V, $V_{IH} = -1.21$ V and $V_{IL} = -1.43$ V.

The noise margins are calculated from Eq. (6-29) as

$$NM_H = V_{OH} - V_{IH} = -0.90 + 1.21 = 0.31 \text{ V}$$

$$NM_L = V_{IL} - V_{OL} = -1.43 + 1.74 = 0.31 \text{ V}$$

Note that the symmetry of the transfer function gives rise to equal noise margins, $NM_H = NM_L = 0.31$ V.[1]

---

[1]Use of the unity-slope points to evaluate $NM_H$ and $NM_L$ yields numerical results that differ by only several millivolts.

**TABLE 6-4 Comparison of Logic Families**

| Family series | TTL | | | CMOS* | | ECL | |
|---|---|---|---|---|---|---|---|
| *Parameter* | *74 LS* | *74 AS* | *74 ALS* | *74 C* | *74 HC* | *10K* | *100K* |
| Nominal supply voltage, V | 5 | 5 | 5 | 5 | 5 | −5.2 | −4.5 |
| Maximun $V_{OL}$, V | 0.5 | 0.5 | 0.5 | 0.4 | 0.4 | −1.7 | −1.7 |
| Minimum $V_{OH}$, V | 2.7 | 2.7 | 2.7 | 4.2 | 4.2 | −0.9 | −0.9 |
| Maximum $V_{IL}$, V | 0.8 | 0.8 | 0.8 | 1.0 | 1.0 | −1.4 | −1.4 |
| Minimum $V_{IH}$, V | 2.0 | 2.0 | 2.0 | 3.5 | 3.5 | −1.2 | −1.2 |
| $NM_H$, V | 0.7 | 0.7 | 0.7 | 0.7 | 0.7 | 0.3 | 0.3 |
| $NM_L$, V | 0.3 | 0.3 | 0.3 | 0.6 | 0.6 | 0.3 | 0.3 |
| Logic swing, V | 2.0 | 2.0 | 2.0 | 3.8 | 3.8 | 0.8 | 0.8 |
| Power dissipation per gate, mW | 2 | 20 | 1 | ≈0 | ≈0 | 24 | 40 |
| Delay-power product, pJ | 20 | 30 | 4 | 30‡ | 10‡ | 48 | 30 |
| Fan-out | 100 | 10 | 100 | >100 | >100 | 10 | 10 |

*Measured at a load current $I_{OL}$ = 4 mA. At $I_{OL}$ = 0.2 mA, $V_{OL}$ = 0.1 V and $V_{OH}$ = 4.8 V. ‡ Delay in ns.

component density on a chip of any technology. Thus it is interesting to note that each of the four major logic families described in this chapter is "best" with respect to one major performance criterion.

## REFERENCES


1 Hodges, D. A., and H. G. Jackson: "Analysis and Design of Digital Integrated Circuits," McGraw-Hill Book Company, New York, 1983.

2 Taub, H., and D. Schilling: "Digital Integrated Electronics," McGraw-Hill Book Company, New York, 1977.

3 Sedra, A. S., and K. C. Smith: "Microelectronic Circuits," Holt, New York, 1982.

4 Ghausi, M. S.: "Electronic Devices and Circuits: Discrete and Integrated," Holt, New York, 1985.

5 Elmasry, M. I. (ed.): "Digital MOS Integrated Circuits," IEEE Press, New York, 1981.

6 Solomon, P. M.: A Comparison of Semiconductor Devices for High-Speed Logic, *Proceedings of IEEE*, vol. 70, no. 5, pp. 489–509, May 1982.

7 *IEEE Journal of Solid-State Circuits*, special issues on semiconductor logic and memory, October 1970 to present.


## REVIEW QUESTIONS

**6-1** What is meant by a binary number?

**6-2** Define (*a*) positive logic; (*b*) negative logic.

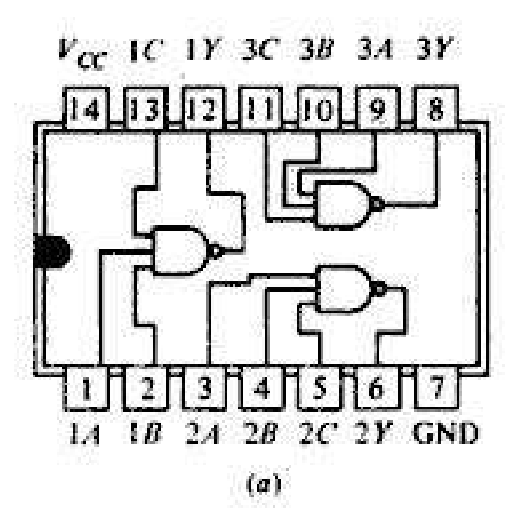
**6-3** What is meant by dynamic logic?

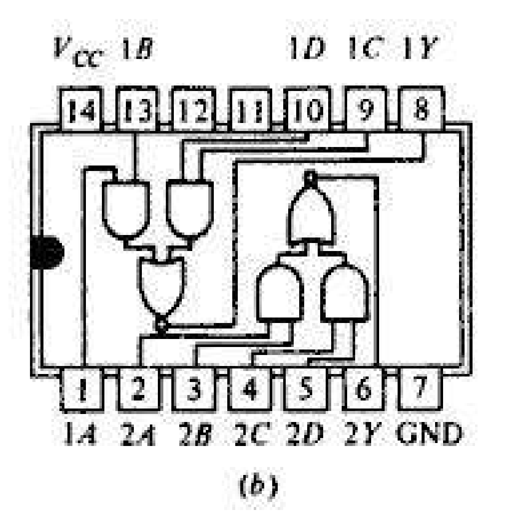
Quad 2-input NAND
Triple 3-input NAND
Dual 4-input NAND
Single 8-input NAND
Dual 4-input AND
Triple 3-input AND
Hex inverter
Dual 2-wide, 2-input AOI
Single 2-wide, 4-input AOI
Single, 4-wide, 4-2-3-2-input AOI
Single, 4-wide, 2-input AOI
Single 4-wide, 2-2-3-2 input AOI

Quad 2-input NOR
Quad 2-input OR
Quad 2-input exclusive-NOR
Quad 2-input exclusive-OR
Triple 3-input NOR
Triple 3-input NOR
Dual 4-input NOR
Triple 3-input OR
Single 8-input NOR
Hex 2-input OR
Triple 4-input OR
Triple 4-input NOR

These combinations are available in most logic families (TTL, CMOS, etc.) listed in Sec. 6-15. The limitation on the number of gates per chip is usually set by the number of pins available. The most common package is the *dual-in-line* (plastic or ceramic) package (DIP), which has 14 leads, 7 brought out to each side of the IC (Fig. 7-1*c*). The dimensions of the assembly, which is much larger than the chip size, are approximately 2 by 0.75 by 0.5 cm. A schematic of the triple 3-input NAND is shown in Fig. 7-1*a*. Note that there are 3 × 3 = 9 input leads, 3 output leads, a power-supply lead, and a ground lead; a total of 14 leads is used.

In Fig. 7-1*b* is indicated the dual 2-wide, 2-input AOI (AND-OR INVERT). This combination needs 4 input leads and 1 output lead per AOI, or 10 for the dual array. If 1 power-supply lead and 1 ground lead are added, we see that 12 of



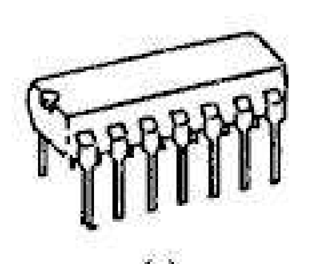




**FIGURE 7-1**
The load connections (top view) of (*a*) the triple 3-input NAND gate. (*b*) A dual 2-wide, 2-input, AND-OR-invert gate. (*c*) A dual-in-line package (DIP).

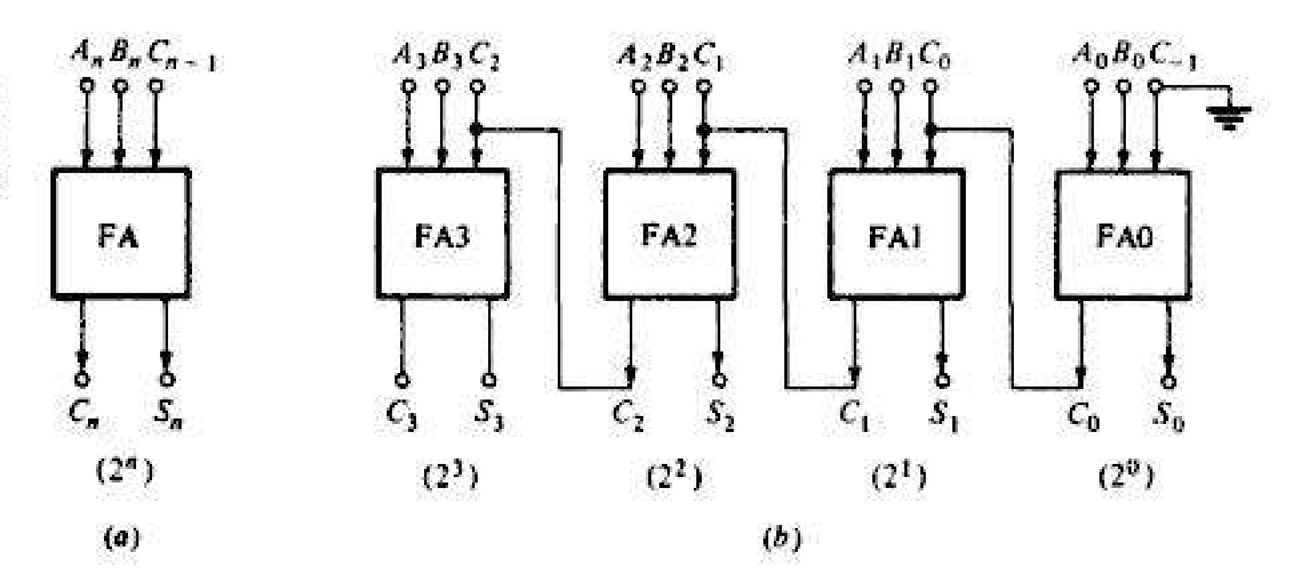



**FIGURE 7-5** (*a*) The symbol for a full adder. (*b*) A 4-bit parallel binary adder using cascaded full adders.

10," then in the last row $D = 0$. Because a 1 must now be carried to the place of next higher significance, $C = 1$.

From Fig. 7-4*b* we see that $D$ obeys the exclusive-OR function and $C$ follows the logic of an AND gate. These functions are indicated in Fig. 7-4*c*, and may be implemented in many different ways with the circuitry discussed in Chap. 6. For example, the exclusive-OR gate can be constructed with any of the four topologies of Sec. 6-15 and in any of the logic families in Table 6-4. The configuration in Fig. 6-8*b* ($Y = A\overline{B} + B\overline{A}$) is implemented in TTL logic with the AOI circuit of Fig. 7-2. The inverter for $B$ (or $A$) is a single-input NAND gate. Since $Y$ has an AND-OR (rather than an AND-OR-INVERT) topology, a transistor inverter is placed between node $P$ and the base of $Q4$ in Fig. 7-2.

**Parallel Operation** Two multidigit numbers may be added serially (one column at a time) or in parallel (all columns simultaneously). Consider parallel operation first. For an $N$-digit binary number there are (in addition to a common ground) $N$ signal leads in the computer for each number. The $n$th line for number $A$ (or $B$) is excited by $A_n$ (or $B_n$), the bit for the $2^n$ digit ($n = 0, 1, \ldots, N - 1$).

**Full Adder** In IC implementation, addition is performed using a complete adder, which (for reasons of economy of components) is not constructed from two half-adders. The symbol for the $n$th full adder (FA) is indicated in Fig. 7-5*a*. The circuit has three inputs: the addend $A_n$, the augend $B_n$, and the input carry $C_{n-1}$ (from the next lower bit). The outputs are the sum $S_n$ (sometimes designated $\Sigma_n$) and the output carry $C_n$. A parallel 4-bit adder is indicated in Fig. 7-5*b*. Since FA0 represents the least-significant bit (LSB), it has no input carry; hence $C_{-1} = 0$.

The circuitry within the block FA may be determined from Fig. 7-6, which is the truth table for adding 3 binary bits. From this table we can verify that the boolean expressions for $S_n$ and $C_n$ are given by

$$S_n = \overline{A}_n\overline{B}_nC_{n-1} + \overline{A}_nB_n\overline{C}_{n-1} + A_n\overline{B}_n\overline{C}_{n-1} + A_nB_nC_{n-1} \qquad (7\text{-}1)$$

$$C_n = \overline{A}_nB_nC_{n-1} + A_n\overline{B}_nC_{n-1} + A_nB_n\overline{C}_{n-1} + A_nB_nC_{n-1} \qquad (7\text{-}2)$$

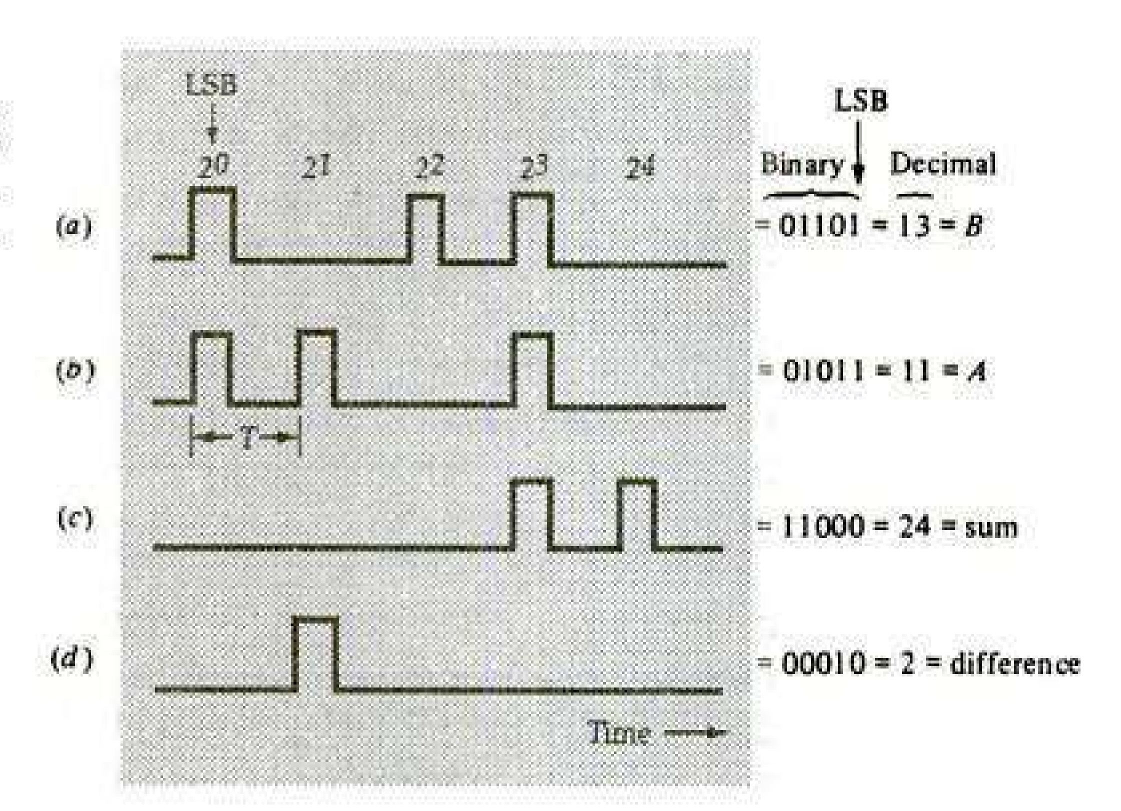



**FIGURE 7-9**
(*a*, *b*) Pulse waveforms representing numbers *B* and *A*. (*c*, *d*) Waveforms for sum and difference.

(24) and difference (2) are shown in Fig. 7-9*c* and *d*, respectively. A serial *adder* is a device which will take as inputs the two waveforms of Fig. 7-9*a* and *b* and deliver the output waveform in Fig. 7-9*c*. Similarly, a *subtractor* (Sec. 7-3) will yield the output shown in Fig. 7-9*d*.

We have already emphasized that the sum of two multidigit numbers may be formed by adding to the sum of the digits of like significance the carry (if any) which may have resulted from the next lower place. With respect to the pulse trains of Fig. 7-9, the above statement is equivalent to saying that, at any instant of time, we must add (in binary form) to the pulses *A* and *B* the carry pulse (if any) which comes from the resultant formed one period *T* earlier. The logic outlined above is performed by the full-adder circuit of Fig. 7-10. This circuit differs from the configuration in the parallel adder of Fig. 7-5 by the inclusion of a time delay TD which is equal to the time *T* between pulses. Hence the carry pulse is delayed a time *T* and added to the digit pulses in *A* and *B*, exactly as it should be.

A comparison of Figs. 7-5 and 7-10 indicates that parallel addition is faster than serial because all digits are added simultaneously in the former, but in sequence in the latter. However, whereas only one full adder is needed for serial arithmetic, we must use a full adder for each bit in parallel addition. Hence parallel addition is much more expensive than serial operation.

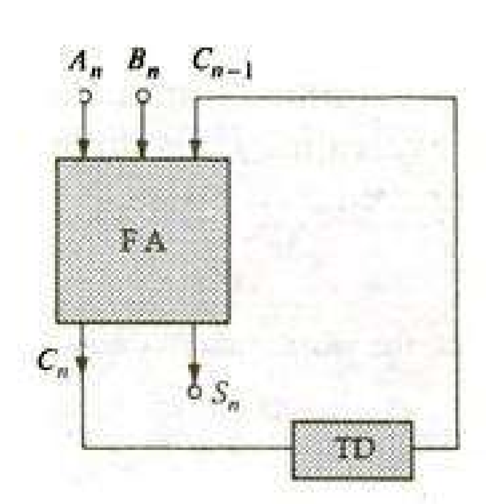
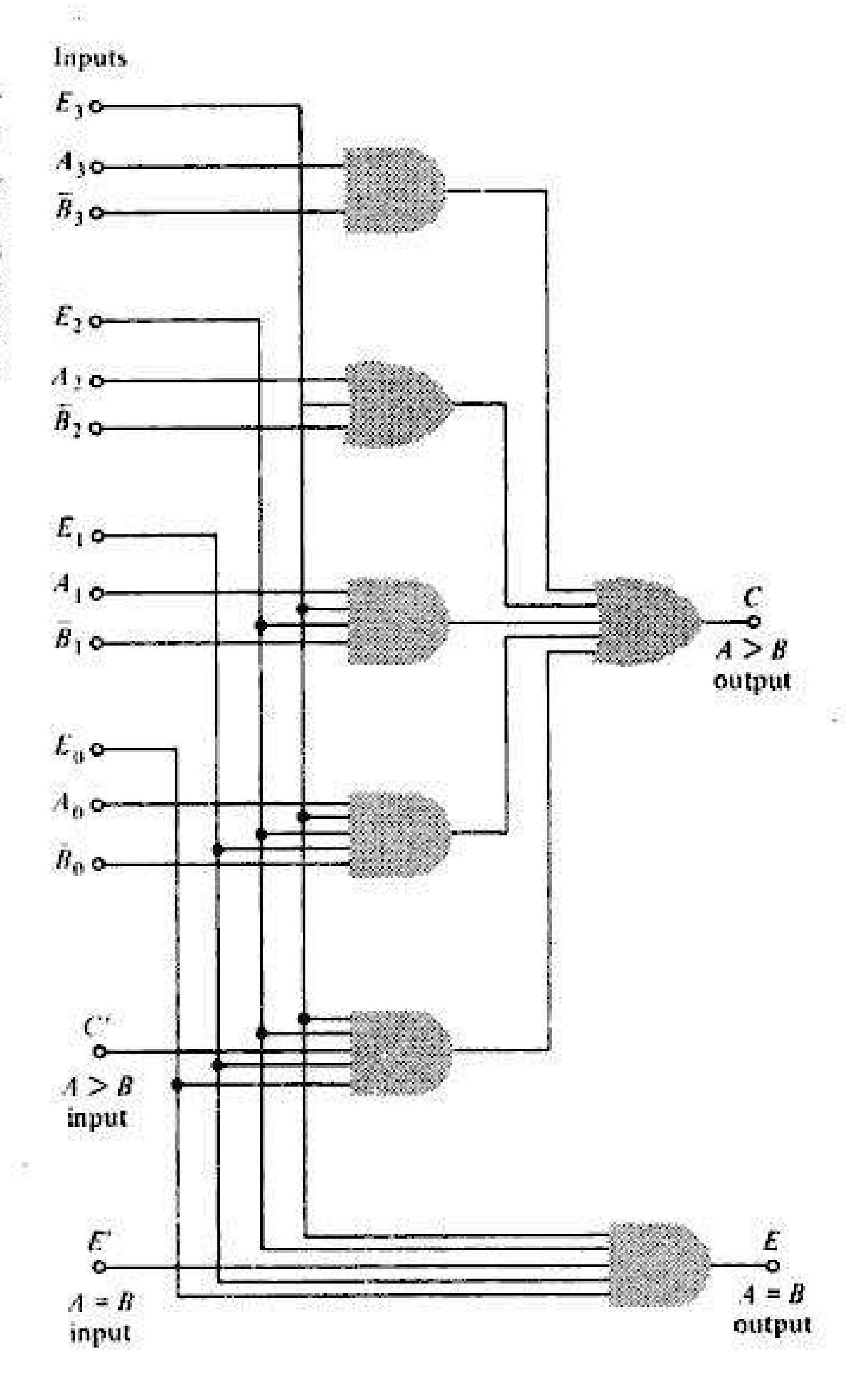



**FIGURE 7-10**
A serial full adder.

**FIGURE 7-13** A 4-bit magnitude comparator. Assume $C' = 0$ and $E' = 1$. If $E = 1$, then $A = B$ and if $C = 1$, $A > B$. If $D = 1$, then $A < B$, where $D$ has the same topology as $C$ but with $A$ and $B$ interchanged.




Consider now a 4-bit comparator. $A = B$ requires that

$$A_3 = B_3 \quad \text{and} \quad A_2 = B_2 \quad \text{and} \quad A_1 = B_1 \quad \text{and} \quad A_0 = B_0$$

Hence the AND gate $E$ in Fig. 7-13 described by

$$E = E_3E_2E_1E_0 \tag{7-12}$$

implies $A = B$ if $E = 1$ and $A \neq B$ if $E = 0$. (Assume that the input $E'$ is held high; $E' = 1$.)

The inequality $A > B$ requires that

$A_3 > B_3$ (MSB)

or $A_3 = B_3$ and $A_2 > B_2$

or $A_3 = B_3$ and $A_2 = B_2$ and $A_1 > B_1$

or $A_3 = B_3$ and $A_2 = B_2$ and $A_1 = B_1$ and $A_0 > B_0$

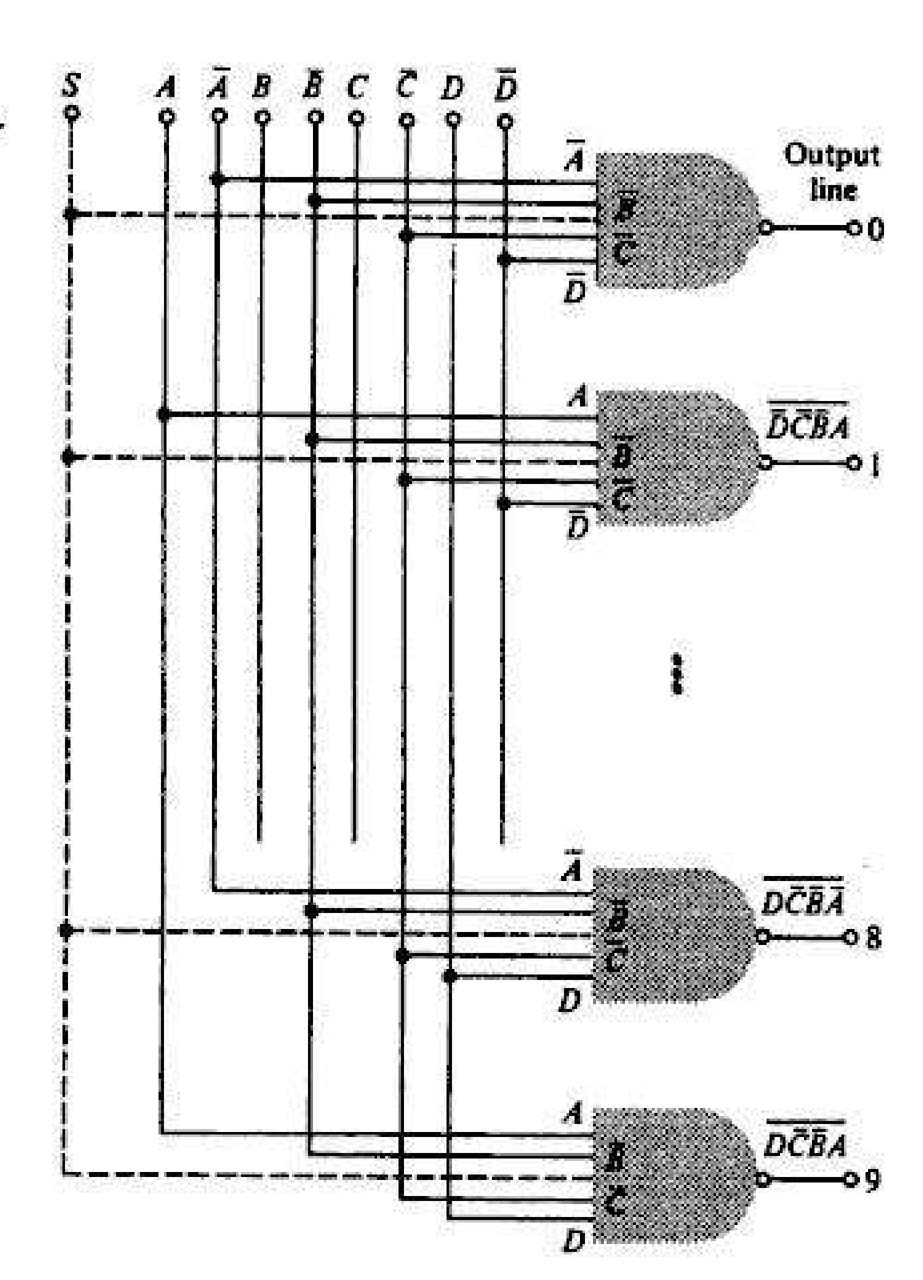



**FIGURE 7-17**
A BCD-to-decimal decoder.

In other words, the decoder acts as a 10-position switch which responds to a BCD input instruction.

It is sometimes desired to decode only during certain intervals of time. In such applications an additional input, called a *strobe*, is added to each NAND gate. All strobe inputs are tied together and are excited by a binary signal $S$, as indicated by the dashed lines in Fig. 7-17. If $S = 1$, a gate is *enabled* and decoding takes place, whereas if $S = 0$, no coincidence is possible and decoding is inhibited. The strobe input can be used with a decoder having any number of inputs or outputs.

**Demultiplexer**

A *demultiplexer* is a system for transmitting a binary signal (serial data) on one of $N$ lines, the particular line being selected by means of an address. A single-pole $N$-position rotary switch connected as in Fig. 7-18*a* is the mechanical analog of such a demultiplexer. The address determines the angle of rotation of the arm of the switch. A decoder is converted into a demultiplexer by means of the dashed connections in Fig. 7-17. If the data signal is applied at $S$, then the output will be the complement of this signal (because the output is 0 if all inputs are 1) and will appear only on the addressed line.

An enabling signal may be applied to a demultiplexer by cascading the system of Fig. 7-17 with that indicated in Fig. 7-19. If the *enable* input is 0, then $S$ is

which represents serially the binary data applied in parallel at the input. In other words, a parallel-to-serial conversion is accomplished of one 16-bit word. This process takes $16T$ seconds.

In a digital system, such as a computer, a data communication system, etc., a pulse train is often required for testing or control (gating) purposes. Such a *sequence generator* is obtained by means of the parallel-to-series converter. Any desired waveform may be obtained by properly choosing the input data $X$.

**Sequential Data Selection** By changing the address with a counter in the manner indicated in the preceding paragraph, the operation of an electromechanical stepping switch is simulated. If the data inputs are pulse trains, this information will appear sequentially on the output channel: in other words, pulse train $X_0$ will appear for $T$ seconds, followed by $X_1$ for the next $T$ seconds, etc. If the number of data sources is $M$, then $X_0$ is again selected during the interval $MT < t < (M + 1)T$.

**Higher-Order Multiplexers** If the number of input lines exceeds 16, then the logic block diagram assumes a topology which is the inverse to that shown in Fig. 7-20. For example, to select 1 out of 32 data inputs, the system in Fig. 7-22 may be used. Multiplexer $N_{20}$ places the data inputs $X_0$ through $X_7$ in sequence onto line $L_0$ as the address

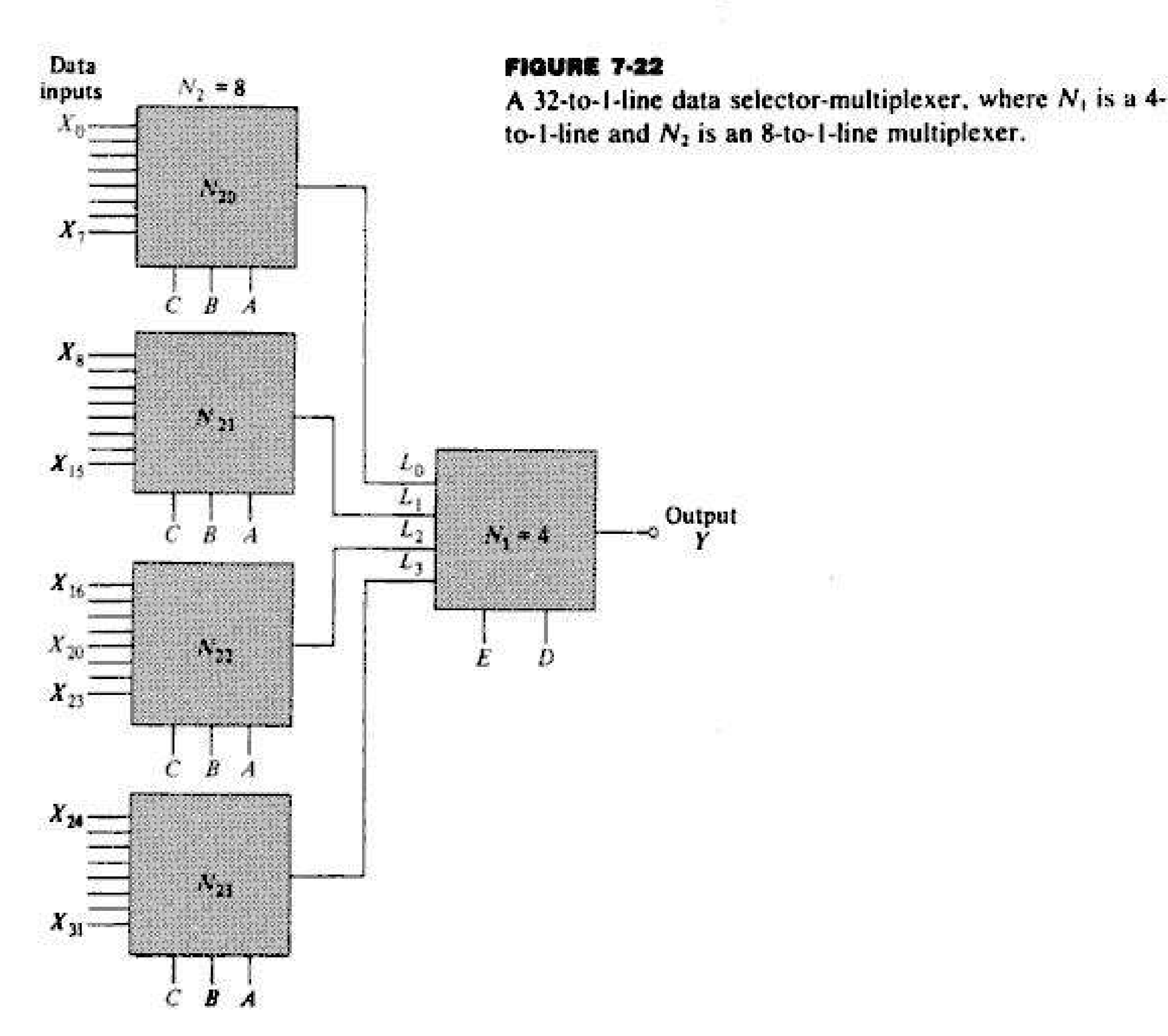



**FIGURE 7-22**
A 32-to-1-line data selector-multiplexer, where $N_1$ is a 4-to-1-line and $N_2$ is an 8-to-1-line multiplexer.

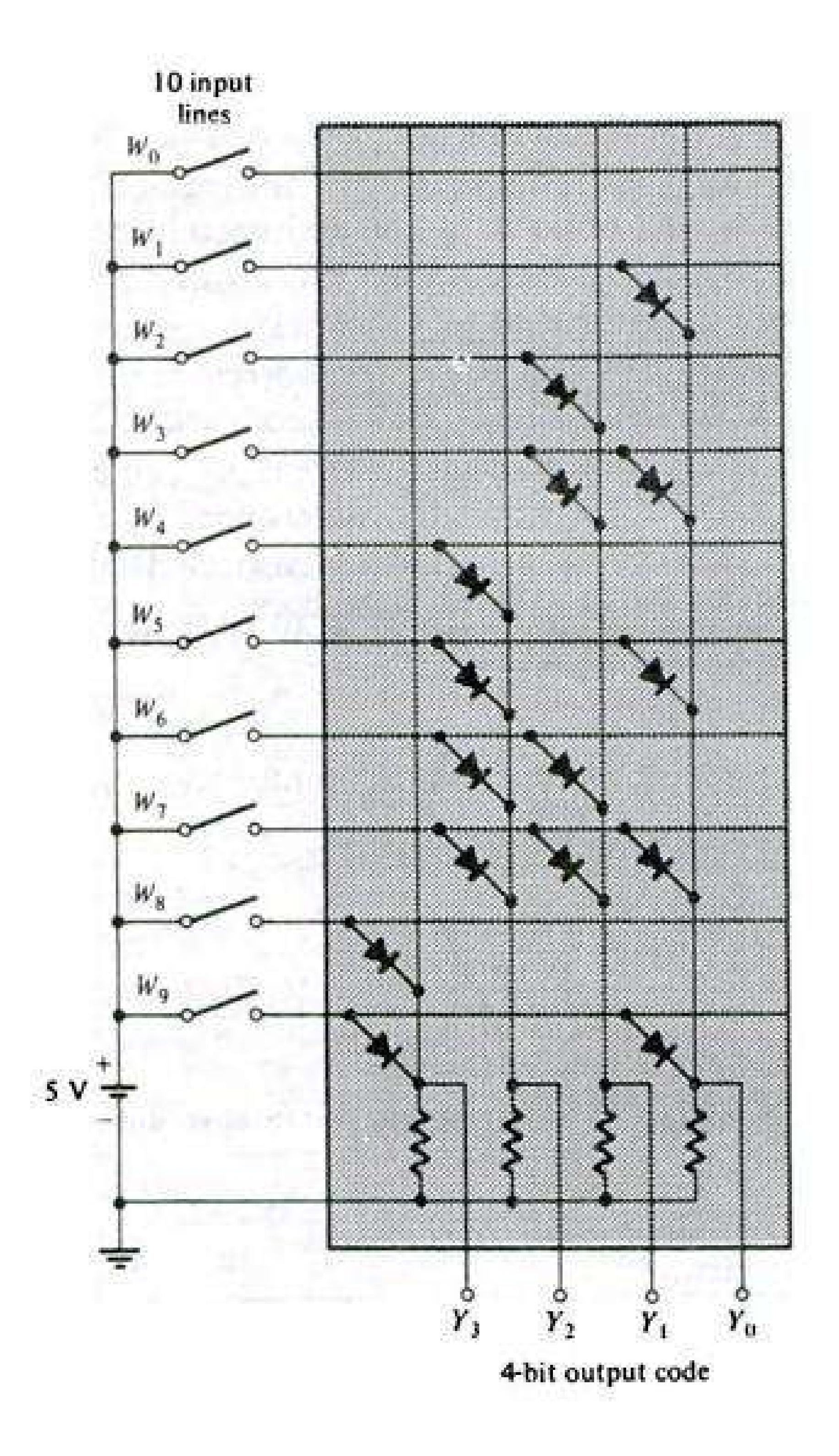
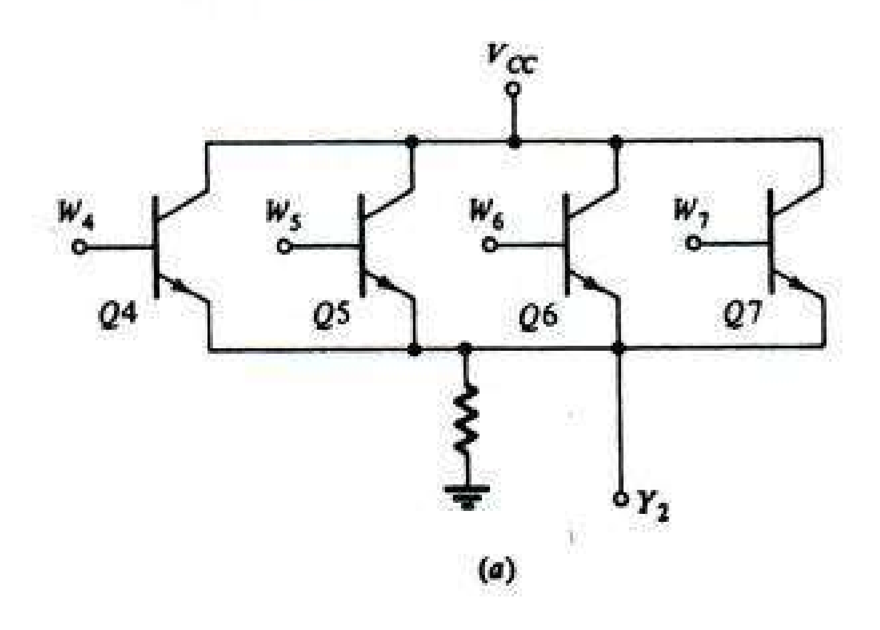
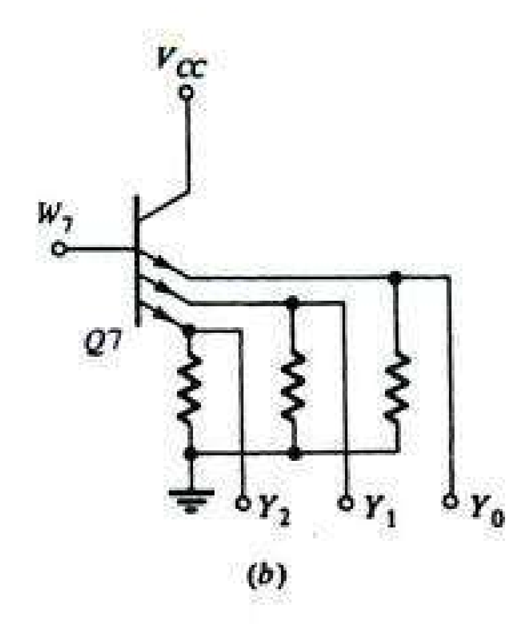



**FIGURE 7-24**
An encoding matrix to transform a decimal number into a BCD code. The key $W_0$ may be omitted since it is implied that the output is 0000 unless one of the other nine keys is activated.

**FIGURE 7-25**
(*a*) An emitter-follower OR gate, (*b*) the line $W_7$ in the encoder of Fig. 7-24 is connected to the base of the three-emitter transistor.



You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

For example

$$W_0 = \bar{X}_3\bar{X}_2\bar{X}_1\bar{X}_0 \qquad W_5 = \bar{X}_3 X_2\bar{X}_1 X_0 \qquad W_9 = X_3\bar{X}_2\bar{X}_1 X_0 \tag{7-26}$$

From the truth table (Table 7-4), we conclude that

$$Y_0 = W_1 + W_2 + W_5 + W_6 + W_9 + W_{10} + W_{13} + W_{14} \tag{7-27}$$

This equation is implemented by connecting eight diodes with their cathodes all tied to $Y_0$ and their anodes connected to the decoder lines $W_1$, $W_2$, $W_5$, $W_6$, $W_9$, $W_{10}$, $W_{13}$, and $W_{14}$, respectively (or the base-emitter diodes of transistors may be used in an analogous manner to form an emitter-follower OR gate, as in Fig. 7-25*a*). Similarly, from Table 7-4, we may write the boolean expressions for the other output bits. For example,

$$Y_3 = W_8 + W_9 + W_{10} + W_{11} + W_{12} + W_{13} + W_{14} + W_{15} \tag{7-28}$$

Consider the inverse code translation, from Gray to binary. The Gray-code inputs [(1) in Table 7-4] are arranged in the order $W_0, W_1, \ldots, W_{15}$ (corresponding to decimal numbers 0 to 15). The binary code corresponding to a given input word $W_n$ is listed as the output code [(3) in Table 7-4] for that line. For example, from (1) and (2) of Table 7-4 at line $W_{14}$, we find that the Gray code 1001 corresponds to the binary code 1110, and this relationship is maintained in Table 7-4 [(1) and (3)] on line $W_9$. From this table we obtain the relationship between binary output (3) and Gray input (1) bits. For example,

$$Y_0 = W_1 + W_2 + W_4 + W_7 + W_8 + W_{11} + W_{13} + W_{14} \tag{7-29}$$

This equation defines how the memory elements are to be arranged in the encoder. Note that the ROM for conversion from a binary to a Gray code uses the same decoding arrangement as that for conversion from a Gray to a binary code. However, the encoders are completely different. In other words, the IC chips for these two ROMs are quite distinct since individual masks must be used for the encoder matrix of memory elements.

**Programming the ROM** Consider a 256-bit bipolar ROM 7488A arranged in 32 words of 8 bits each. The decoder input is a 5-bit binary select code, and its outputs are the 32 word lines. The encoder consists of 32 transistors (each base is tied to a different line) and with 8 emitters in each transistor. The customer fills out the truth table he wishes the ROM to satisfy, and then the vendor makes a mask for the metallization so as to connect one emitter of each transistor to the proper output line, or alternatively, to leave it floating. For example, for the Gray-to-binary-code conversion, Eq. (7-29) indicates that one emitter from each of transistors *Q*1, *Q*2, *Q*4, *Q*7, *Q*8, *Q*11, *Q*13, and *Q*14 is connected to line $Y_0$, whereas the corresponding emitter on each of the other transistors *Q*0, *Q*3, *Q*5, *Q*6 . . . is left unconnected. The process just described is called *custom programming* or *mask programming* of a ROM. Note that *hardware* (not *software*) programming is under consideration.

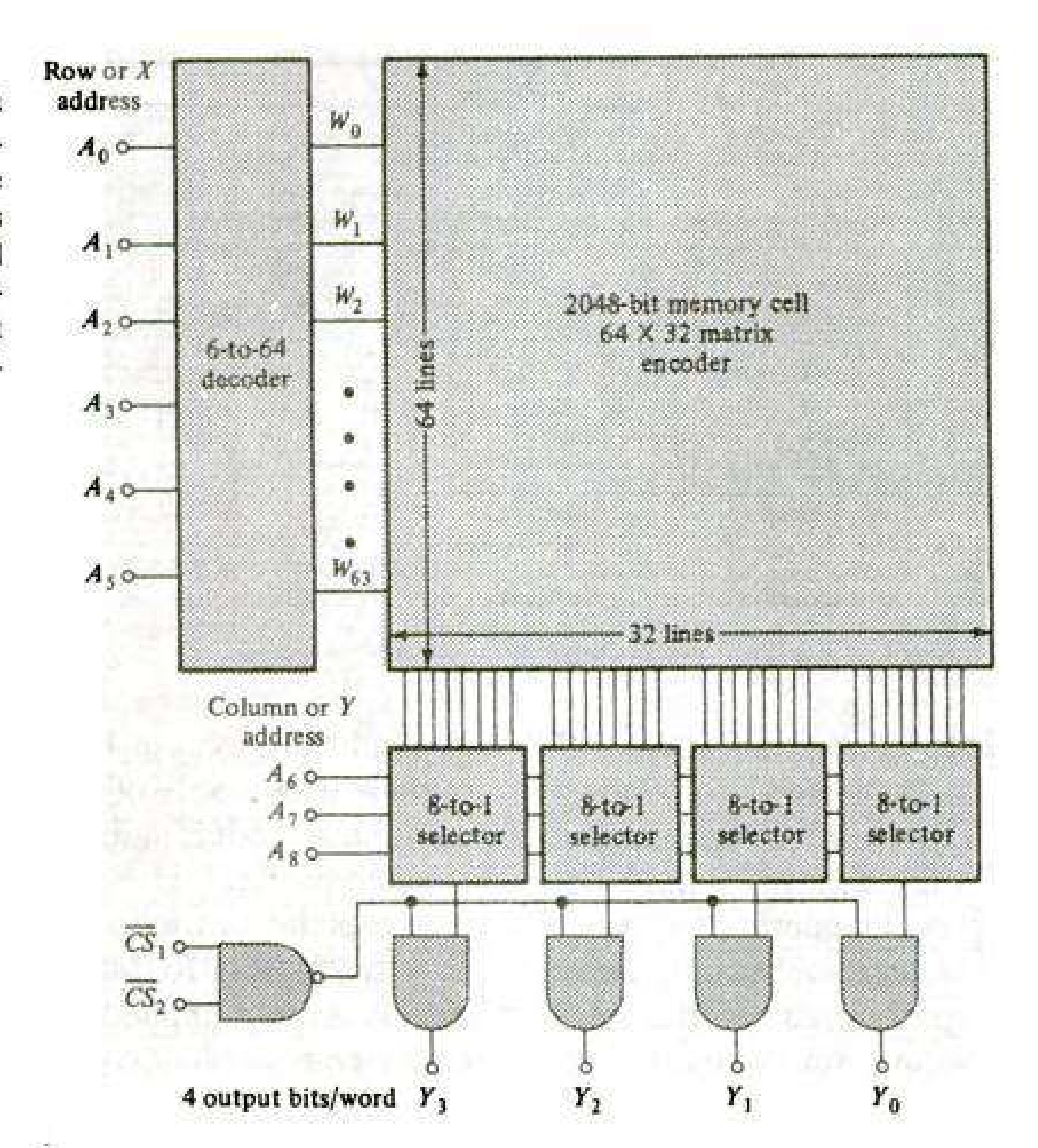



**FIGURE 7-30**
A 2-kb ROM (512 × 4 bits) with two-dimensional addressing. Note that the column address $A_8A_7A_6$ is applied to all four selectors-multiplexers. The chip-select input $CS$ is used for enabling purposes.

9 and $N = 4$). A total of 512 NAND gates are required in the decoder, one for each word line. Considerable economy results if the topology shown in Fig. 7-30 is used. A 6-bit $X$ (row) address generates 64 horizontal lines. If 32 $Y$ (vertical) lines are used in the memory matrix, the total number of bits is $64 \times 32 = 2048$ as required. However, since only four output lines are specified then four 8-to-1-line selectors are used. A 3-bit column address feeds each multiplexer. This arrangement is called $X$-$Y$ or *two-dimensional addressing*. Note that now 64 NAND gates are needed for the decoder and $4 \times 9 = 36$ for the selectors from the NAND-NAND (AND-OR) configuration in Fig. 7-21 (it is clear that an 8-input selector requires nine gates). The total of $64 + 36 = 100$ NANDS for $X$-$Y$ addressing is a tremendous saving over 512 NANDS required by the decoder arrangement of Fig. 7-26 for the same-size ROM. From Fig. 7-30, 64 transistors with 32 emitters each are needed for the encoder whereas in Fig. 7-26 there are 512 transistors, each with 4 emitters.

**Word Expansion** Increasing the number of bits per word is easily accomplished. For example, a 512×8-bit ROM is obtained by using two 512 × 4 ROMs. The addressing in Fig. 7-27 is applied to both chips simultaneously. The 4 bits of lower significance are obtained from one package and the 4 bits of higher significance are taken from the second chip.

$$Y_0 = \bar{D}\bar{C}\bar{B}A + \bar{D}C\bar{B}\bar{A} + \bar{D}CB\bar{A} + D\bar{C}B\bar{A} + D\bar{C}BA + DC\bar{B}\bar{A} + DCB\bar{A} + DCBA \quad (7\text{-}33)$$

There are a number of algebraic and graphical techniques and computer programs for minimizing such boolean expressions. Note, for example, that the second and third minterms can be simplified to

$$\bar{D}C\bar{B}\bar{A} + \bar{D}CB\bar{A} = \bar{D}C\bar{A}$$

because $\bar{B} + B = 1$. Proceeding in this manner (Prob. 7-40), the following minimized form of $Y_0$ is obtained:

$$Y_0 = \bar{D}\bar{C}\bar{B}A + C\bar{A} + DB \quad (7\text{-}34)$$

Using the minimized expressions for $Y_0, Y_1, \ldots, Y_6$ results in some saving (about 20 percent) of components over those required in the ROM. A chip fabricated in this manner (e.g., 74HC4511) is designated a "BCD-to-seven-segment decoder/driver."

Minimization of boolean equations (particularly if the number of variables in each product exceeds five) is tedious and time-consuming. The engineering man-hours cost for minimization and for designing the special IC chip to realize the savings in components must be compared with that of simply programming an existing ROM. Unless a tremendous number of units are to be manufactured (and particularly if the matrix size is large), the ROM is the more economical procedure. The programmable logic arrays discussed in the next section afford a convenient method for designing complex logic functions by use of "fixed" hardware.

**Combinational Logic** If $N$ logic equations of $M$ variables are given in the sum-of-products canonical form, these equations may be implemented with an $M$-input, $N$-output ROM. As explained above, this is an economical solution if $M$ and $N$ are large (particularly if $M$ is large). However, in the logic design of one stage of a full adder, where $M = 3$ and $N = 2$ (small numbers), and where this unit is sold in considerable quantities, using distinct gate combinations as in Fig. 7-7 is more economical than using a ROM.

**Character Generator** Alphanumeric characters may be "written" on the face of a cathode-ray tube (a television-type display) with the aid of a ROM.

**Stored Programs** Control programs (e.g., in a pocket calculator) are permanently stored in ROM.

## 7-12 PROGRAMMABLE ROMS (PROMS)

Many manufacturers supply *field-programmable* ROMs, called *PROMs* (see App. B-1). These IC chips can provide flexibility to the designer and can often reduce costs, particularly when only a small quantity of a particular ROM is required. The cost of making the connection

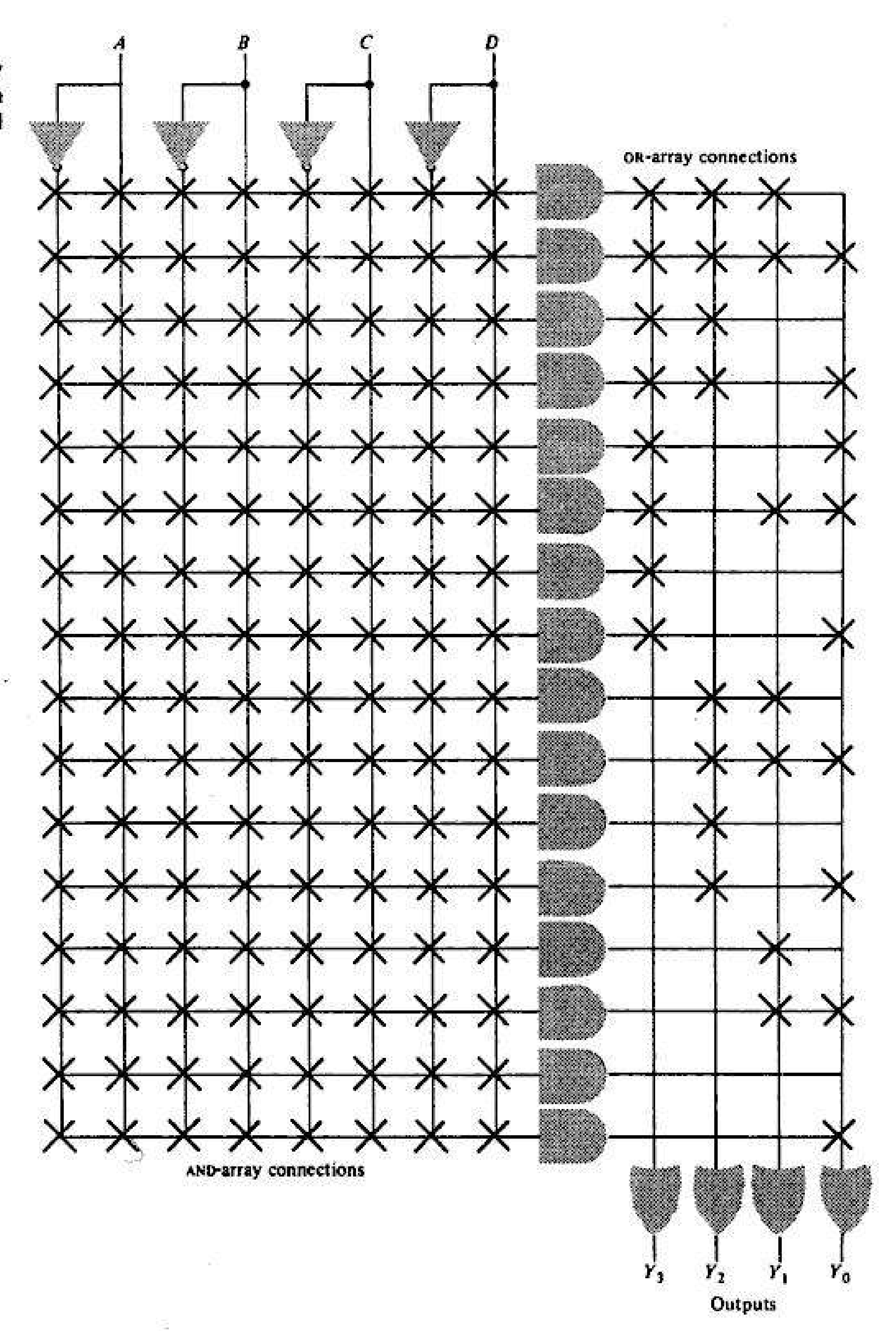



**FIGURE 7-35**
Programmable array logic (PAL) formed from a programmable AND and fixed OR array.

In many logic designs, simplified boolean expressions are implemented by the PAL arrangement compared with PROM realizations. A 16-word × 4-bit PAL is illustrated in Fig. 7-35, in which the X's indicate connections (fusible links).

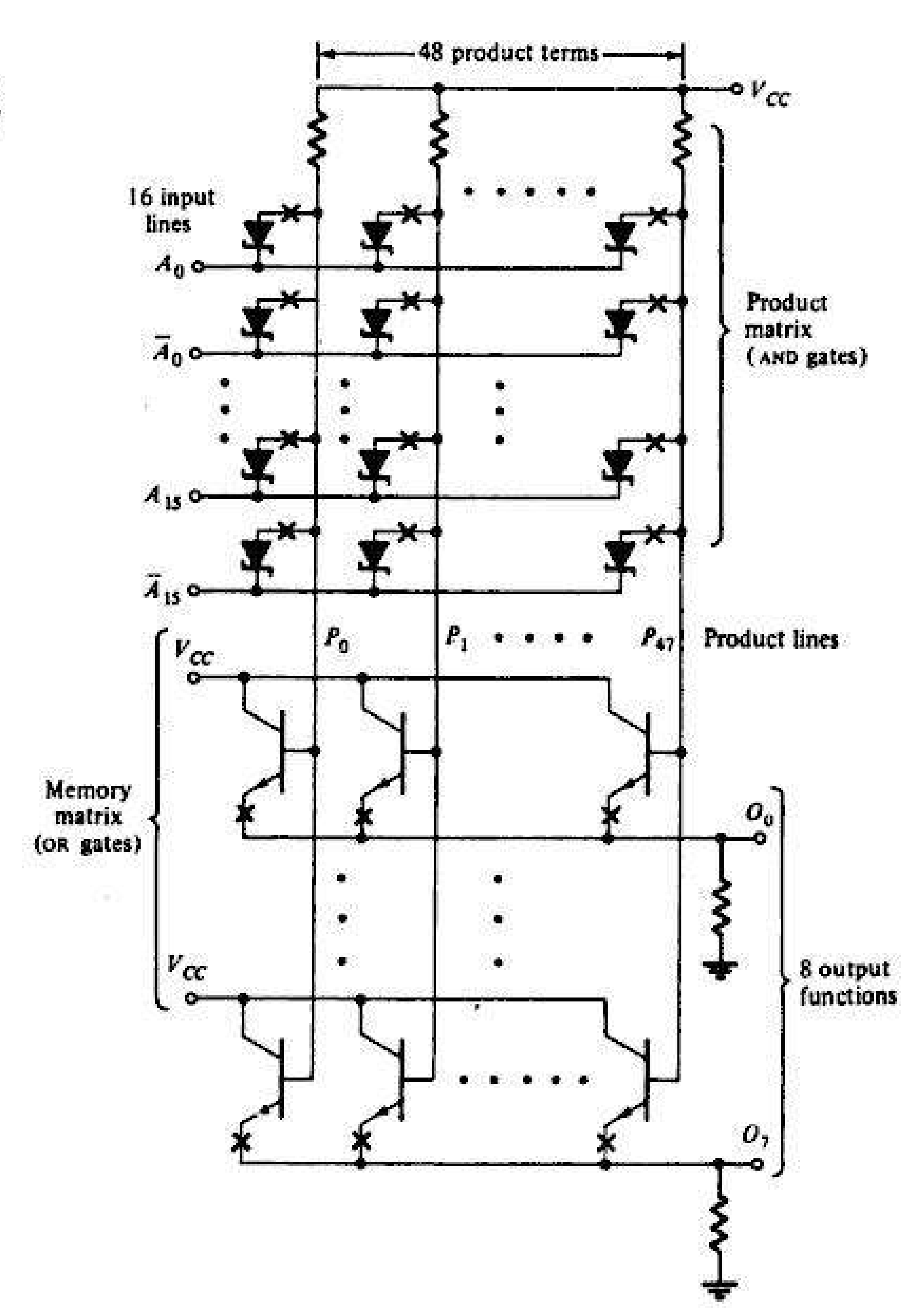



**FIGURE 7-37** The 82S100 FPLA (16 × 48 × 8). (*Courtesy of Signetics.*)

represents a fusible link. This system is programmed in the field by selectively blowing fuses so as to break the connections required in order to satisfy the program table, as explained in the preceding paragraph. The complement $\bar{A}_k$ of $A_k$ is obtained by an inverter (not shown in Fig. 7-37). The complement of each output is also available on the chip, but this circuit is omitted from the figure for simplicity. The package is TTL compatible with either tristate or open-collector outputs (Fig. 6-39) and also includes chip-enable control.

The PLA (or FPLA) is designed for the implementation of complex logic functions. A PLA can handle more data inputs and is more economical than a ROM. It is useful for the same type of applications as the ROM (Sec. 7-11), provided that the number of product terms needed is a small fraction of the total input combinations possible.

# Chapter 8
# SEQUENTIAL CIRCUITS AND SYSTEMS

Many digital systems are required to operate in synchronism with a sequence of binary signals (a pulse train). The operation of a digital computer, for example, depends on first obtaining an instruction from memory (the fetch portion of a machine cycle) and storing it in a register until the instruction is executed. Moreover, the data to be processed must be obtained from its memory location. The third step is executing the instruction. Simultaneously, a program counter must be incremented to prepare for the next instruction. The timing and sequence of these steps is critical to successful operation of the system. Sequential circuits and systems are used to process binary signals synchronously. In addition, circuits capable of storing binary signals (memory) are needed.

The basic building block incorporated in sequential circuits is the FLIP-FLOP. In this chapter we treat FLIP-FLOPS and registers and counters, two classes of circuits based on FLIP-FLOPS. Several typical applications of these circuits are also described.

## 8-1 A 1-BIT MEMORY

All the systems discussed in Chap. 7 were based on combinational logic; the outputs at a given instant of time depend only on the values of the inputs at the same moment. Such a system is said to have no memory. Note that a ROM is a combinational circuit and, according to the above definition, it has no memory. *The memory of a ROM refers to the fact that it "memorizes" the functional relationship between the output variables and the input variables.* It does *not* store bits of information.

**A 1-Bit Storage Cell** The basic digital memory circuit is obtained by cross-coupling two NOT circuits $N1$ and $N2$ (single-input NAND gates) in the manner shown in Fig. 8-1*a*. The output of each gate is connected to the input of the other, and this feedback combination is called a *latch*. The most important property of the latch is that it can exist in one of two stable states, either $Q = 1$ ($\overline{Q} = 0$), called the *1 state*, or $Q = 0$ ($\overline{Q} = 1$), referred to as the *0 state*. The existence

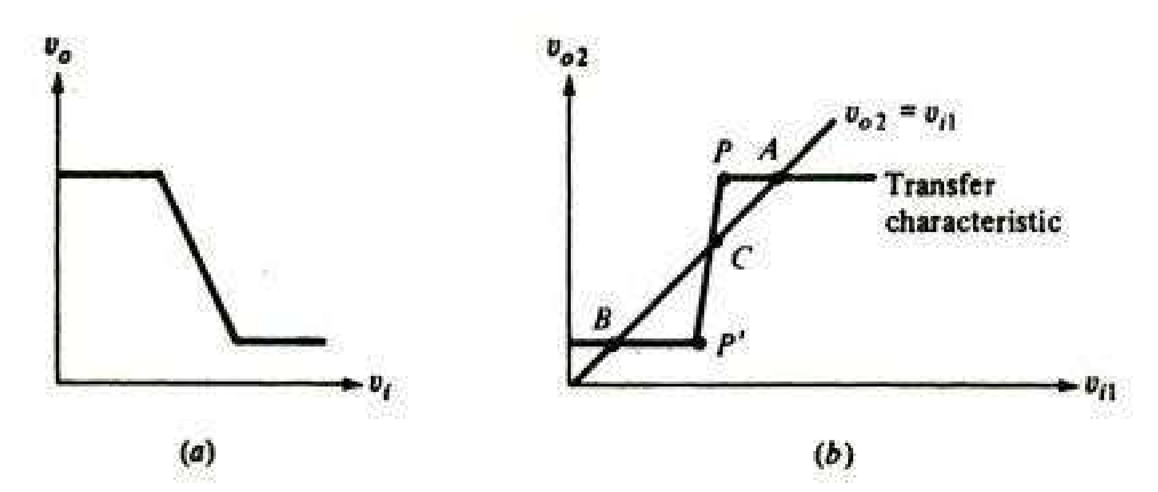



**FIGURE 8-5**
(*a*) Inverter transfer characteristic. (*b*) The voltage transfer characteristic of cascaded inverters (black). The feedback constraint (blue) indicates that points *A* and *B* are the two stable states of the latch. (Point *C* is dynamically unstable.)

with a pulse sequence. The *SR* latch is a common building block in sequential circuits, as is described in detail in Sec. 8-3. The topology of the 74L5279A is displayed in Fig. 8-6, in which the two NAND gates and feedback paths are identified. The ECL latch in Fig. 8-7 uses the NOR outputs of the basic OR/NOR gate. The circuits in Figs. 8-8*a* and 8-8*b* are NMOS and CMOS realizations, respectively, of *SR* latches which employ a coupled NOR-gate topology.

**FIGURE 8-6**
The 74LS5279 *SR* latch constructed from two TTL LS NAND gates.

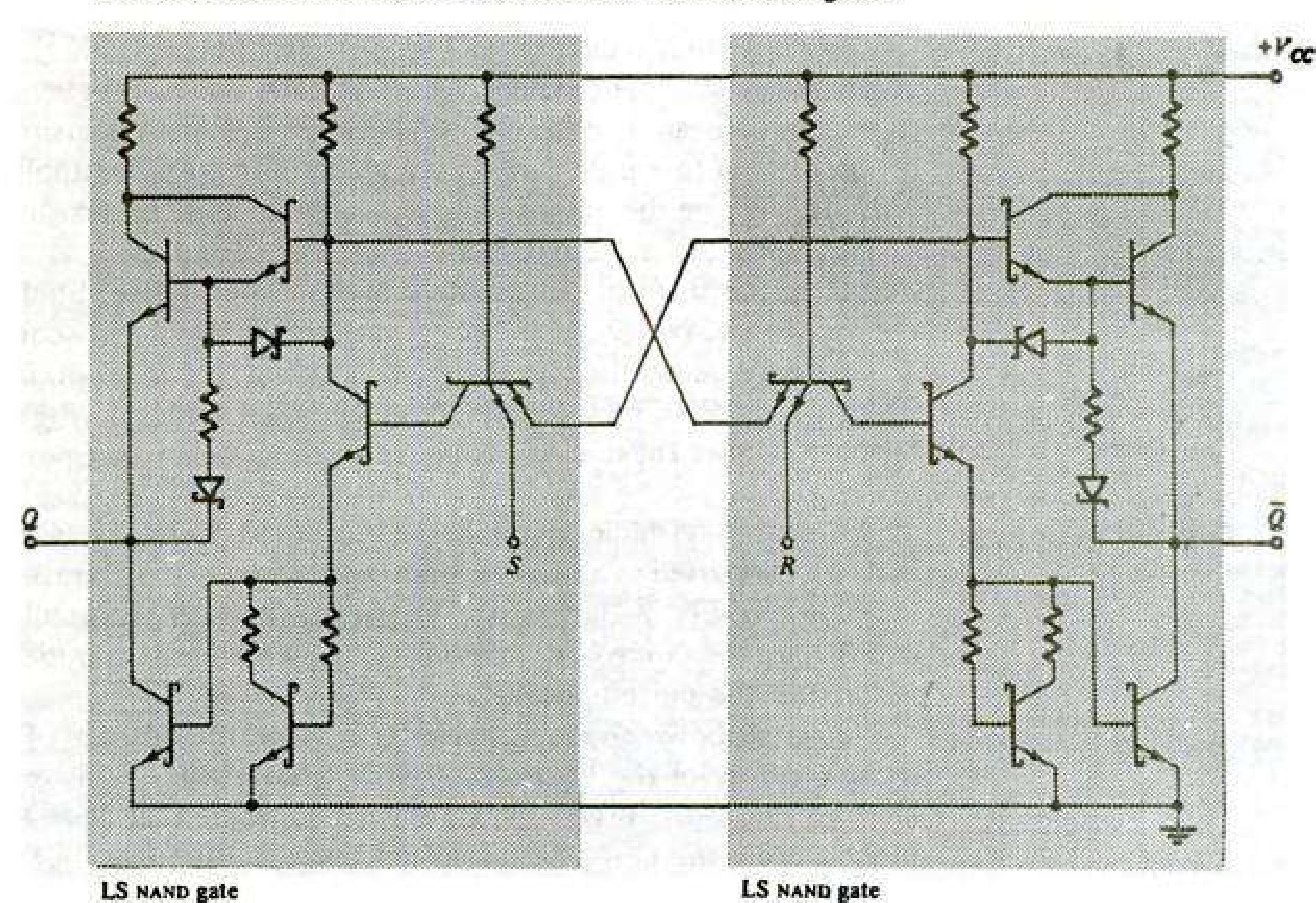
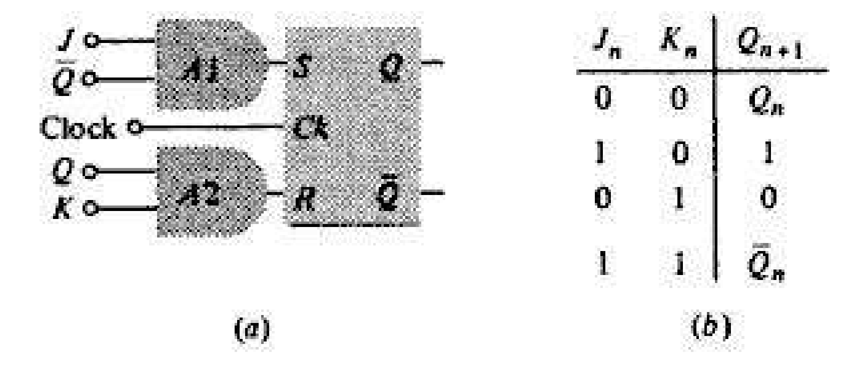


**FIGURE 8-11**
(*a*) Conversion of an *SR* into a *J-K* FLIP-FLOP. (*b*) The truth table.




| $J_n$ | $K_n$ | $Q_{n+1}$ |
|---|---|---|
| 0 | 0 | $Q_n$ |
| 1 | 0 | 1 |
| 0 | 1 | 0 |
| 1 | 1 | $\overline{Q}_n$ |

(*a*) (*b*)

$Q$ are applied to $A2$, and hence $R = KQ$. The logic followed by this system is given in the truth table in Fig. 8-11*b*. This logic can be verified by referring to Table 8-1. There are four possible combinations for the two data inputs $J$ and $K$. For each of these there are two possible states for $Q$, and hence Table 8-1 has eight rows. From the $J_n$, $K_n$, $Q_n$, and $\overline{Q}_n$ bits in each row, $S_n = J_n\overline{Q}_n$ and $R_n = K_nQ_n$ are calculated and are entered into the fifth and sixth columns of the table. Using these values of $S_n$ and $R_n$ and referring to the *SR* FLIP-FLOP truth table of Fig. 8-10*b*, the seventh column is obtained. Finally, column 8 follows from column 7 because $Q_n = 1$ in row 4, $Q_n = 0$ in row 5, $\overline{Q}_n = 1$ in row 7, and $\overline{Q}_n = 0$ in row 8.

Columns 1, 2, and 8 of Table 8-1 form the *J-K* FLIP-FLOP truth table in Fig. 8-11*b*. Note that *the first three rows of a J-K truth table are identical with the corresponding row for an SR truth table* (Fig. 8-5*b*). However, the ambiguity of the state $S_n = 1 = R_n$ is now replaced by $Q_{n+1} = \overline{Q}_n$ for $J_n = 1 = K$. *If the two data inputs in the J-K* FLIP-FLOP *are high, the output will be complemented by the clock pulse.*

It is really not necessary to use the AND gates $A1$ and $A2$ of Fig. 8-11*a*, since the same function can be performed by adding an extra input terminal to each NAND gate $N3$ and $N4$ of Fig. 8-10*a*. This simplification is indicated in Fig. 8-12. (Ignore the dashed inputs; i.e., assume that they are both 1.) Now $Q$ and $\overline{Q}$ at the inputs are obtained by the feedback connections (drawn in color) from the outputs.

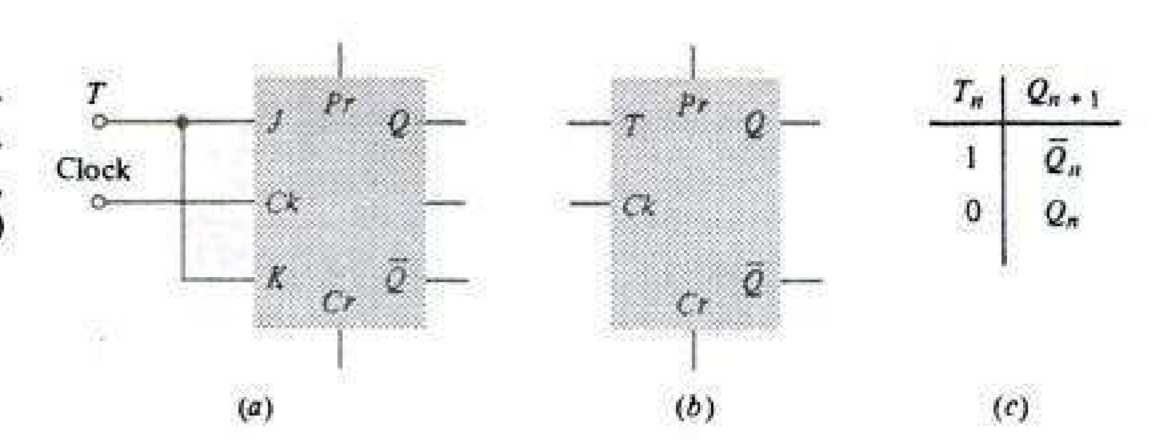
## Preset and Clear

The truth table in Fig. 8-11*b* tells us what happens to the output with the application of a clock pulse, as a function of the data inputs $J$ and $K$. However,

**TABLE 8-1 Truth Table for Fig. 8-11a**

| Column | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|
| Row | $J_n$ | $K_n$ | $Q_n$ | $\overline{Q}_n$ | $S_n$ | $R_n$ | $Q_{n+1}$ | |
| 1 | 0 | 0 | 0 | 1 | 0 | 0 | $Q_n$ | $Q_n$ |
| 2 | 0 | 0 | 1 | 0 | 0 | 0 | $Q_n$ | |
| 3 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 1 |
| 4 | 1 | 0 | 1 | 0 | 0 | 0 | $Q_n$ | |
| 5 | 0 | 1 | 0 | 1 | 0 | 0 | $Q_n$ | 0 |
| 6 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | |
| 7 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | $\overline{Q}_n$ |
| 8 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | |

**FIGURE 8-15**
(*a*) A *J-K* FLIP-FLOP converted into a *T*-type FLIP-FLOP with data input *T*, (*b*) the logic symbol, (*c*) the truth table.




| $T_n$ | $Q_{n+1}$ |
|---|---|
| 1 | $\bar{Q}_n$ |
| 0 | $Q_n$ |

changes only after *Ck* changes from 1 to 0, at the *negative-going* (*trailing*) *edge* of the pulse. It is also possible to design a *J-K* FLIP-FLOP so that the output changes at the *positive-going* (*leading*) *edge* of the pulse. The 74LS109A chip is a dual positive-edge-triggered *J-K* FLIP-FLOP with preset and clear inputs. The MC10H176 contains six (hex) positive-edge-triggered *D*-type FLIP-FLOPS.

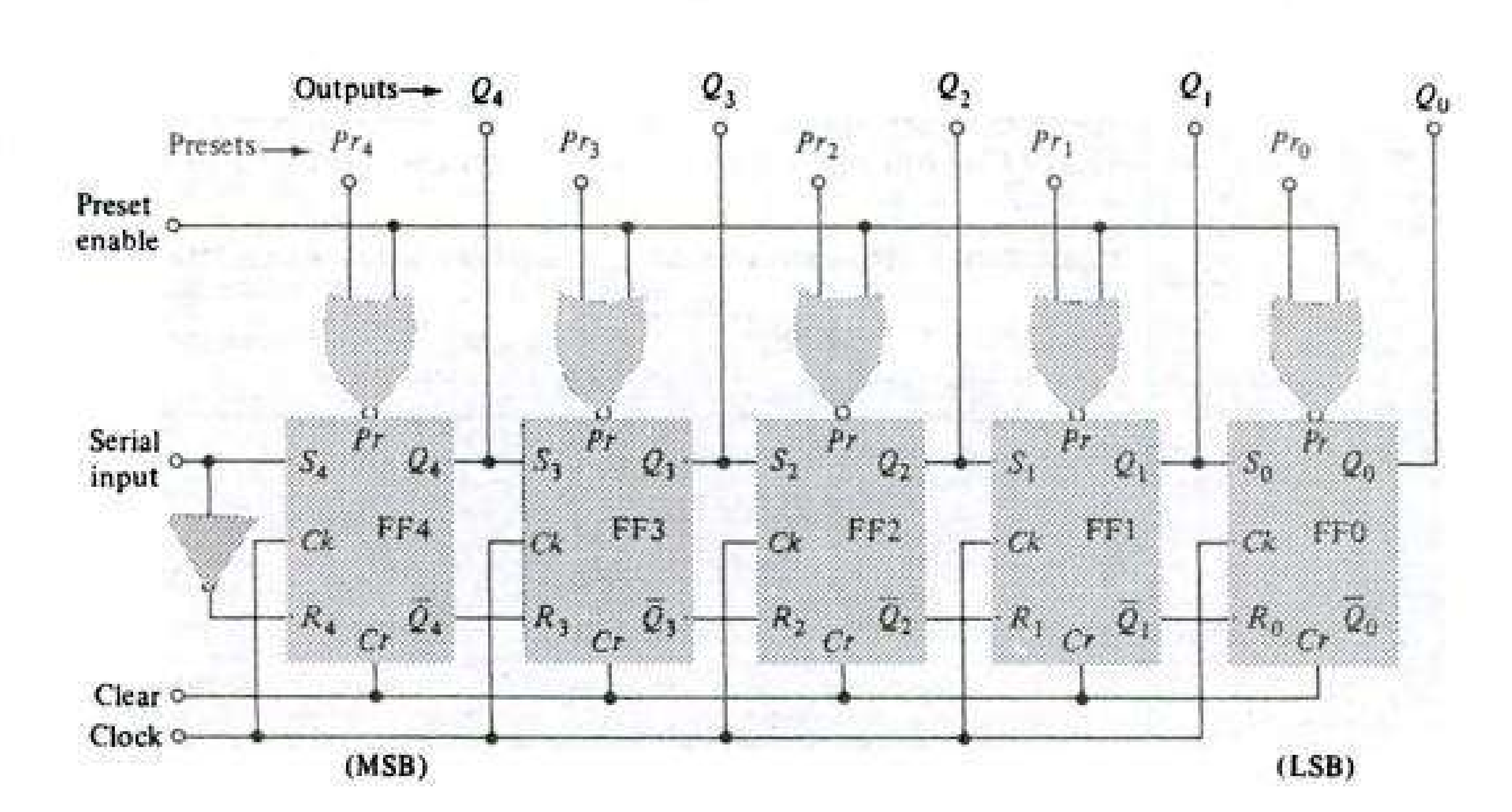
The toggle, or complementing, FLIP-FLOP is not available commercially because a *J-K* can be used as a *T* type by connecting the *J* and *K* inputs together (Fig. 8-15).

The FLIP-FLOP is available in all the IC digital families, and the maximum frequencies of operation are given in Table 6-4.

## 8-5 SHIFT REGISTERS

Since a binary is a 1-bit memory, then *n* FLIP-FLOPS can store an *n*-bit word. This combination is referred to as a *register*. To allow the data in the word to be read into the register serially, the output of one FLIP-FLOP is connected to the input of the following binary. Such a configuration, called a *shift register*, is indicated in Fig. 8-16. Each FLIP-FLOP is of the *SR* (or *J-K*) master-slave type. Note that the stage which is to store the most significant bit (MSB) is converted into a *D*-type latch (Fig. 8-14) by connecting *S* and *R* through an inverter. The 5-bit shift register indicated in Fig. 8-16 is available on a single chip in a 16-pin package (medium-scale integration). We shall now explain the

**FIGURE 8-16**
A 5-bit shift register.



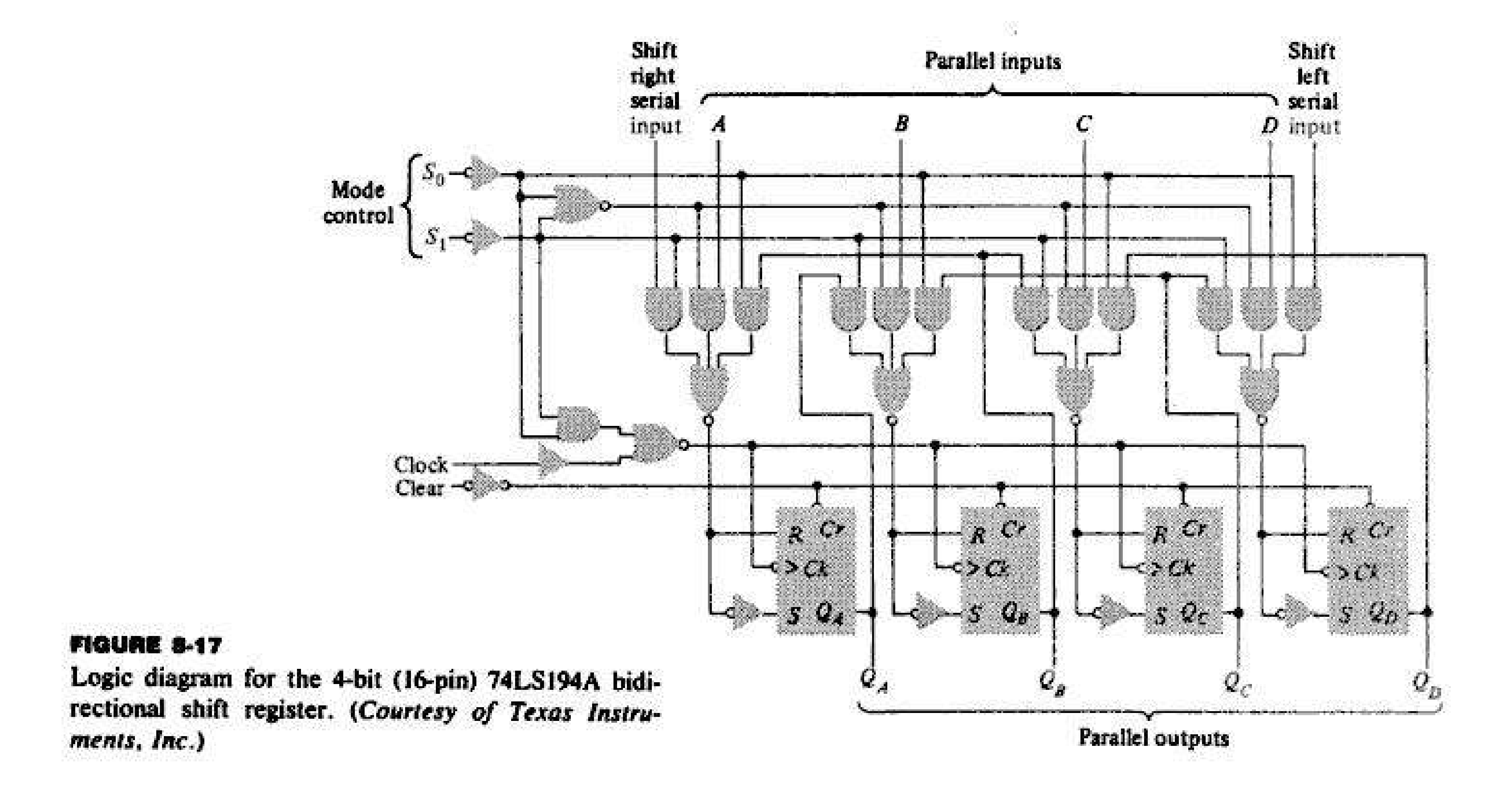



**FIGURE 8-17**
Logic diagram for the 4-bit (16-pin) 74LS194A bidirectional shift register. (*Courtesy of Texas Instruments, Inc.*)

A sequence generator may also be obtained from a multiplexer (Sec. 7-7) and a number of simultaneous sequences may be generated using a ROM (Sec. 7-9).

**Shift-Register Ring Counter** Consider the 5-bit shift register (Fig. 8-16) with $Q_0$ connected to the serial input. Such a circulating memory forms a *ring counter*. Assume that all FLIP-FLOPS are cleared and then that FF0 is preset so that $Q_0 = 1$ and $Q_4 = Q_3 = Q_2 = Q_1 = 0$. The first clock pulse transfers the state of FF0 to FF4, so that after the pulse $Q_4 = 1$ and

$$Q_3 = Q_2 = Q_1 = Q_0 = 0$$

Succeeding pulses will transfer the state 1 progressively around the ring. The count is read by noting which FLIP-FLOP is in state 1; no decoding is necessary.

Consider a ring counter with $N$ stages. If the interval between triggers is $T$, then the output from any binary stage is a pulse train of period $NT$, with each pulse of duration $T$. The output pulse of one stage is delayed by a time $T$ from a pulse in the preceding stage. These pulses may be used where a set of sequential gating waveforms is required. Thus a ring counter is analogous to a stepping switch, where each triggering pulse causes an advance of the switch by one step.

Since there is one output pulse for each $N$ clock pulses, the counter is also a *divide-by-N* unit, or an $N$:1 *scaler*. Typically, TTL shift-register counters operate at frequencies as high as 25 MHz.

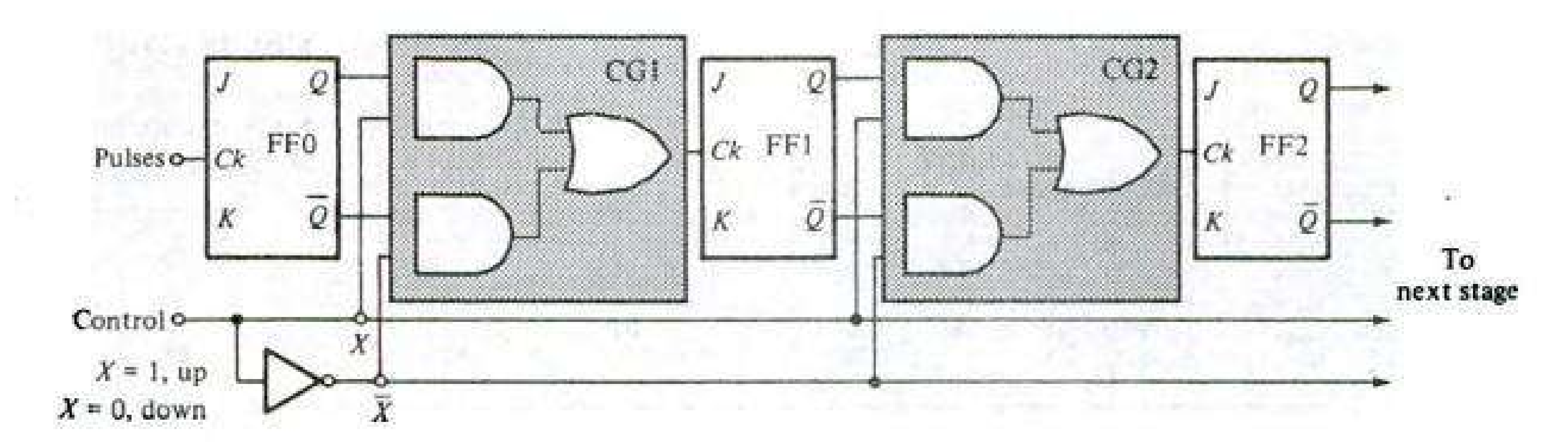



**FIGURE 8-20**
An up-down ripple counter. (It is understood that $J = K = 1$.)

counter it is always to be understood that $J = K = 1$ as in Fig. 8-18. The two-level AND-OR gates CG1 and CG2 between stages constitute a 2-input multiplexer which controls the direction of the counter. Note that this logic combination is equivalent to a NAND-NAND configuration. If the input $X$ is a 1 (0), then $Q$ ($\overline{Q}$) is effectively connected to the following FLIP-FLOP and pulses are added (subtracted). In other words, $X = 1$ converts the system to an *up* counter and $X = 0$ to a *down* counter. The control $X$ may not be changed from 1 to 0 (or 0 to 1) between input pulses, because a spurious count may be introduced by this transition. (The synchronous counter of Fig. 8-22 does not have this difficulty and hence up-down counters are operated synchronously, Sec. 8-7).

**Divide-by-*N* Counter** It may be desired to count to a base $N$ which is not a power of 2. We may prefer, for example, to count to the base 10, since the decimal system is the one with which we are most familiar. To construct such a counter, start with a ripple chain of $n$ FLIP-FLOPS such that $n$ is the smallest number for which $2^n > N$. Add a feedback gate so that at count $N$ all binaries are reset to zero. This feedback circuit is simply a NAND gate whose output feeds all *clear* inputs in parallel. Each input to the NAND gate is a FLIP-FLOP output $Q$ which becomes 1 at the count $N$.

Let us illustrate the above procedure for a decade counter. Since the smallest value of $n$ for which $2^n > 10$ is $n = 4$, then four FLIP-FLOPS are required. The decimal number 10 is the binary number 1010 (LSB), and hence $Q_0 = 0$, $Q_1 = 1$, $Q_2 = 0$, and $Q_3 = 1$. The inputs to the feedback NAND gate are therefore $Q_1$ and $Q_3$, and the complete circuit is shown in Fig. 8-21*a*. Note that after the tenth pulse $Q_1$ and $Q_3$ both go to 1, the output of the NAND gate becomes 0, and all FLIP-FLOPS are cleared (reset to 0). (Note that $Q_1$ and $Q_3$ first become 1 and then return to 0 after pulse 10, thereby generating a narrow spike.)

If the propagation delay from the clear input to the FLIP-FLOP output varies from stage to stage, the clear operation may not be reliable. In the above example, if FF3 takes an appreciably longer time to reset than FF1, then when $Q_1$ returns to 0, the output of the NAND gate goes to 1, so that $Cr = 1$ and $Q_3$ will not reset. Wide variations in reset propagation time may occur if the counter

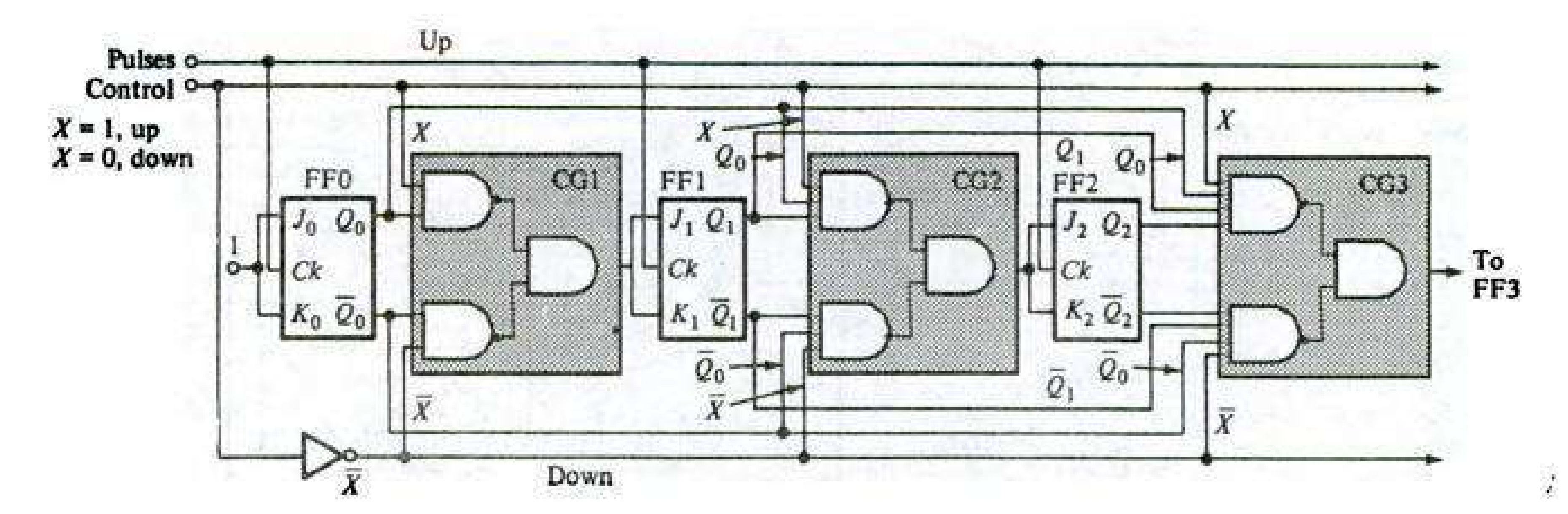



**FIGURE 8-23**
An up-down synchronous counter with parallel carry. The control $X$ may be changed from up to down (or vice versa) between input pulses without introducing spurious counts because the counter responds only to the application of a clock pulse.

With a great deal of patience and intuition, the design may be carried out from direct observation of the waveform chart. Consider, for example, the synthesis of a synchronous decade counter with parallel carry. The waveform chart is that given in Fig. 8-19 except that *after the tenth pulse all waveforms return to* 0. Since $Q_0 = 0$ and $Q_2 = 0$ after the tenth pulse, FF0 and FF2 are excited as in the 16:1 synchronous counter. Hence, from Eq. (8-1)

$$T_0 = J_0 = K_0 = 1 \qquad T_2 = J_2 = K_2 = Q_0Q_1 \tag{8-5}$$

We note from Fig. 8-19 that FF1 toggles if $Q_0 = 1$. However, to prevent $Q_1$ from going to 1 after the tenth pulse, it is inhibited by $Q_3$. These statements are equivalent to the statement

$$T_1 = J_1 = K_1 = Q_0\overline{Q}_3 \tag{8-6}$$

Finally, we wish FF3 to change state from 0 to 1 after the eighth pulse and to return to 0 after the tenth pulse. If

$$J_3 = Q_0Q_1Q_2 \qquad K_3 = Q_0 \tag{8-7}$$

then the desired logic is followed because $Q_0 = Q_1 = Q_2 = 1$, so that $J_3 = 1$, $K_3 = 1$, before pulse 8, whereas $Q_0 = 1$, $Q_1 = 0$, and $Q_2 = 0$, so that $J_3 = 0$, $K_3 = 1$, before pulse 10. The implementation of Eqs. (8-5) to (8-7) is given in the logic block diagrams of Fig. 8-24.

Synchronous up-down decade counters are available commercially (for example, MC10137 or 74ALS168) on a single MSI chip, as are 4-bit binary counters such as MC10154 or 74LS697. The FLIP-FLOPS are provided with *preset* (so that they are programmable) and *clear* inputs, not indicated in Fig. 8-23. Division by a number other than 2, 5, 6, 10, 12, or a power of 2 is not commercially available and must be designed as explained in the foregoing.

**8-8 APPLICATIONS OF COUNTERS** Many systems, including digital computers, data handling, and industrial control systems, use counters. We describe briefly some of the fundamental applications.

**8-3** (*a*) Define a sequential system.
(*b*) How does it differ from a combinational system?

**8-4** What is meant by a stable state?

**8-5** (*a*) Draw the transfer characteristic of a bistable latch.
(*b*) Explain why only two possible states exist in practice.

**8-6** (*a*) Sketch the logic system for a clocked *SR* FLIP-FLOP.
(*b*) Verify that the state of the system does not change in between clock pulses.
(*c*) Give the truth table.
(*d*) Justify the entries in the truth table.

**8-7** (*a*) Augment an *SR* FLIP-FLOP with two AND gates to form a *J-K* FLIP-FLOP.
(*b*) Give the truth table.
(*c*) Verify part *b* by making a table of $J_n$, $K_n$, $Q_n$, $\overline{Q}_n$, $S_n$, $R_n$, and $Q_{n+1}$.

**8-8** Explain what is meant by a race-around condition in connection with the *J-K* FLIP-FLOP in Rev. 8-7.

**8-9** (*a*) Draw a clocked *J-K* FLIP-FLOP system and include *preset* (*Pr*) and *clear* (*Cr*) inputs.
(*b*) Explain the clear operation.

**8-10** (*a*) Draw a master-slave *J-K* FLIP-FLOP system.
(*b*) Explain its operation and show that the race-around condition is eliminated.

**8-11** (*a*) Show how to convert a *J-K* FLIP-FLOP into a delay (*D*-type) unit.
(*b*) Give the truth table.
(*c*) Verify this table.

**8-12** Repeat Rev. 8-11 for a toggle (*T*-type) FLIP-FLOP.

**8-13** Give the truth tables for each FLIP-FLOP type: (*a*) *SR*; (*b*) *J-K*; (*c*) *D*; and (*d*) *T*. What are the direct inputs *Pr* and *Cr* and the clock *Ck* for (*e*) presetting; (*f*) clearing; and (*g*) normal clocked operation?

**8-14** (*a*) Define a register.
(*b*) Construct a shift register from *SR* FLIP-FLOPS.
(*c*) Explain its operation.

**8-15** (*a*) Explain why there may be a race condition in a shift register.
(*b*) How is this difficulty bypassed?

**8-16** Explain how a shift register is used as a converter from (*a*) serial-to-parallel data and (*b*) parallel-to-serial data.

**8-17** Explain how a shift register is used as a sequence generator.

**8-18** Explain how a shift register is used as a circulating read-only memory.
(*a*) Explain how a shift register is used as a ring counter.
(*b*) Draw the output waveform from each FLIP-FLOP of a three-stage unit.

**8-19** (*a*) Sketch the block diagram for a Johnson (twisted-ring) counter.
(*b*) Draw the output waveform from each FLIP-FLOP of a three-stage unit.
(*c*) By what number *N* does this system divide?

**8-20** (*a*) Draw the block diagram of a ripple counter.
(*b*) Sketch the waveform at the output of each FLIP-FLOP for a three-stage counter.
(*c*) Explain how this waveform chart is obtained.
(*d*) By what number *N* does this system divide?

1. The MOS is a bidirectional switch.
2. The very high input resistance permits temporary data storage on the small gate-to-substrate capacitance of a MOS device.
3. The load FET may be turned off by a clock pulse to reduce static power dissipation.

The inverter discussed above is called a *ratioed inverter*. The name derives from the fact that when the input is high and the clock is high, transistors $Q1$ and $Q2$ form a voltage divider between $V_{DD}$ and ground. Therefore, the output voltage $V_o$ depends on the ratio of the ON resistance of $Q1$ and the effective load resistance of $Q2$ (typically, $< 1{:}5$). This ratio is related to the aspect ratios of $Q1$ and $Q2$.

**Two-Phase Ratioed Memory Cell** Cascading two of the dynamic inverters in Fig. 9-1 allows each bit of information which is stored on the capacitance $C$ of the first NOT gate to be transferred to the following inverter by applying a second clock pulse out of phase with the first waveform. A typical MOS dynamic-shift-register stage is shown in Fig. 9-2*a* and the required two-phase clock waveforms are indicated in Fig. 9-2*b*. The waveforms in Fig. 9-2*b* show a *nonoverlapping* two-phase clock as $t_3 > t_2$. When $t_3 < t_2$, the result is an *overlapping clock*. Each stage of the register requires six MOSFETs. The input $V_i$ is the voltage on the gate capacitance $C_1$ of $Q1$, applied there by the previous stage (or by the input signal if this is the first stage of the shift register). When at $t = t_1$ the clock $\phi_1$ goes positive (for NMOS devices), transistors $Q1$ and $Q2$ form an inverter and the bidirectional switch $Q3$ conducts. Hence, the complement of the level of $C_1$ is transferred to $C_2$. When $\phi_1$ drops to 0 (at $t = t_2^+$), $Q2$ and $Q3$ are OFF and $C_2$ retains its charge as long as $\phi_1$ remains at 0 V. However, at $t = t_3^+$, when $\phi_2 = V_{DD}$, then $Q4$ and $Q5$ act as an inverter and the switch $Q6$ is closed.

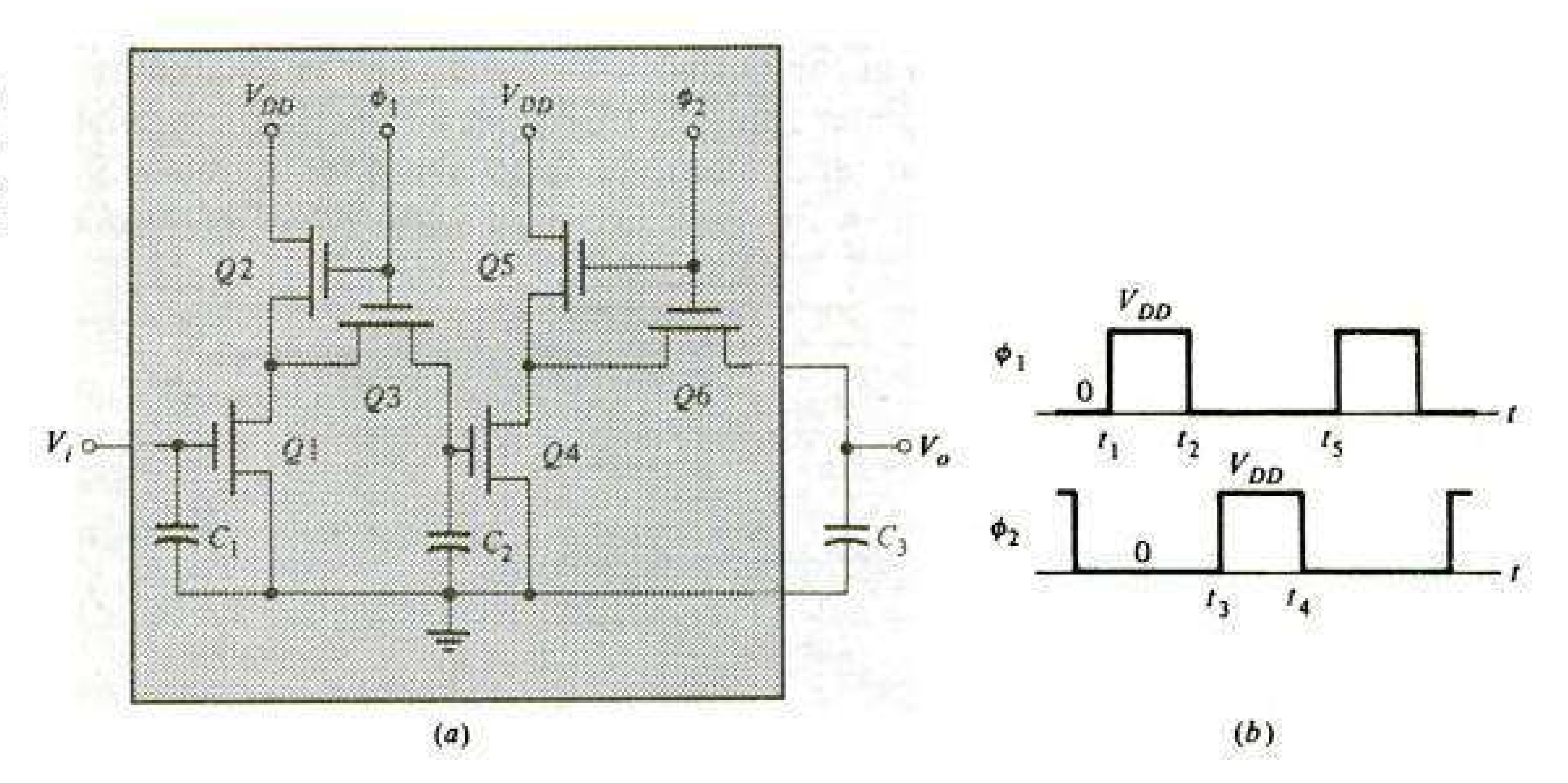



**FIGURE 9-2**
(*a*) A two-phase ratioed dynamic NMOS shift-register stage. (*b*) The two-phase clock waveforms $\phi_1$ and $\phi_2$.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

Since $\phi_1$ is now at its low level, $Q0$ opens and $V_i$ is retained on $C_0$ until the end of the period of $\phi_1$ (at $t = t_5$).

At $t = t_3^+$ the second waveform $\phi_2$ goes to its high level $V_{DD}$ allowing transmission through $Q3$ and effectively placing $C_1$ and $C_2$ in parallel. If at $t = t_3^-$ the voltage on $C_1(C_2)$ is $V_1(V_2)$, then at $t = t_3^+$ the voltage $V$ on $C_2$ (which must be the same as that on $C_1$) is found in Prob. 9-5 to be

$$V = \frac{C_1V_1 + C_2V_2}{C_1 + C_2} \tag{9-1}$$

If $C_1 \gg C_2$, note from Eq. (9-1) that $V \approx V_1$. In other words, pulse $\phi_2$ causes the output voltage (across $C_1$) of the first inverter to appear at the input (across $C_2$) of the second NOT gate. Finally, by the inverter action described above, at the end of the pulse $\phi_2$ (at $t = t_4^+$ and until $t = t_5$) the logic level $V_0$ across $C_3$ is the complement of that across $C_2$, which, in turn, is the complement of that across $C_0$. Clearly, in one period of the clock the input level $V_i$ has shifted through the stage to the output $V_o$, as it should in a 1-bit delay line or 1-bit shift register.

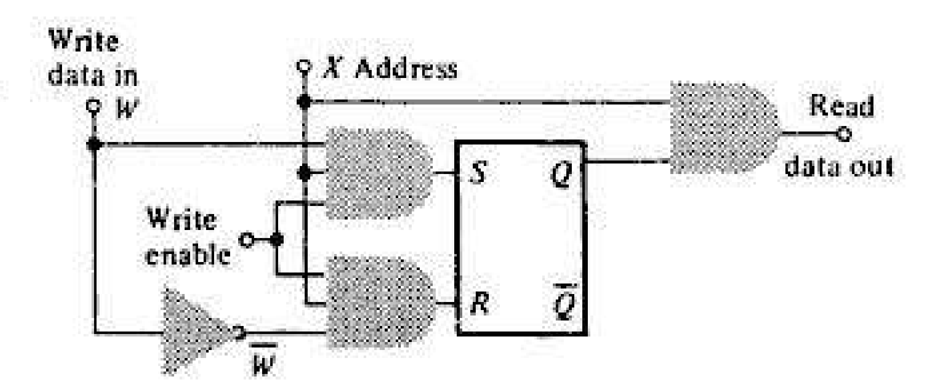
No dc power supply is used in Fig. 9-5, but the clocking waveforms must be capable of furnishing the heavy capacitive currents. Also in order to ensure that $C_1$ be much larger than $C_2$, additional area must be added to the chip for $C_1$. We can reduce the loading on the clock drivers by adding another transistor to each inverter as in Prob. 9-6. This modification results in an eight-MOSFET stage. A number of four-phase ratioless shift registers capable of operation at high speed are described in the literature.[1] Because of the large amount of chip area required for two-phase ratioless shift registers and the additional complications of four-phase clock drivers these systems are seldom used.

**A Dynamic CMOS Shift-Register Stage** A dynamic CMOS shift-register stage, similar to the NMOS circuit in Fig. 9-5, can be constructed by interposing bidirectional CMOS transmission gates (Sec. 6-9) between CMOS static inverters (Sec. 6-8). Such a circuit, depicted in Fig. 9-6, utilizes the transmission gates $T1$ and $T2$ to fulfill the function of the MOS bidirectional switches in Fig. 9-5. The transmission gates are controlled by the complementary clocks $\phi$ and $\overline{\phi}$. When $\phi = V_{DD}$, $T1$ conducts whereas $T2$ acts as an open circuit. The CMOS inverters are marked $I1$ and $I2$.

The explanation of the operation of the register stage of Fig. 9-6 closely parallels that given in connection with Fig. 9-5. When $\phi = V_{DD}$ (logic 1), then $T1$ transmits and the input $V_i$ appears across $C_0$. Because of the inverter action of $I1$, the complement of $V_i$ appears across $C_1$ ($V_1 = \overline{V}_i$). On the next half-cycle $\phi = 0$, $T1$ opens, $C_0$ retains the voltage $V_i$, and $V_1$ remains at the $\overline{V}_i$. Also when $\phi = 0$, $T2$ closes, putting $C_2$ in parallel with $C_1$, and $I2$ causes the voltage across $C_3$ to be the complement of that across $C_2$. Consequently, at the end of a complete cycle $V_o = \overline{V}_1 = V_i$ and we have demonstrated that this cell behaves as a 1-bit delay line or register.

[1] See Refs. 1, 2, 4, and 5 at the end of this chapter.

**FIGURE 9-8**
A 1-bit read/write memory.




Suppose that we wish to have a 16-kb RAM organized as 1024 words of 16 bits each. This system requires 10 addresses, 16 data in and 16 data out lines. A total of 1024 × 16 = 16,384 storage cells must be used. Of this number, 16 cells are arranged in a horizontal line, all excited by the same address line. There are 1024 such lines, each excited by a different address. In other words, addressing is provided by exciting 1 of 1024 lines. This type of addressing is called *one-dimensional* or *linear selection* (Prob. 9-10). The number of pins on the package for addressing is reduced from the unreasonable number of 1024 to only 10 by including on the chip a 10-to-1024-line decoder.

**Two-Dimensional Addressing** A great economy of the number of NAND gates needed in the decoder mentioned above can be obtained (Prob. 9-11) by arranging the 1024 memory elements in a rectangular 32 × 32 array, each cell storing 1 bit of one word. Sixteen such packages are required, one for each of the 16 bits in each word.

Each word is identified by a matrix number $X$-$Y$ in a memory cell of the rectangular matrix. To read (or write into) a specific cell (say, 1-3), an $X$ decoder identifies row 1 ($X_1$) and a $Y$ decoder locates column 3 ($Y_3$). Such *two-dimensional addressing* (also called *X-Y addressing* or *X-Y selection*) is indicated in Fig. 9-9 for a 16-kb (128 × 128) RAM.

**Basic RAM Organization** In the 1-bit memory of Fig. 9-8 separate read and write leads are required. For either the bipolar or the MOS RAM it is possible to construct a FLIP-FLOP (as we demonstrate in Figs. 9-12 and 9-14) which has a common terminal for both writing and reading, such as terminals 1 and 2 in Fig. 9-10. This configuration requires the use not only of the write data $W$ (write 1), but also of its complement $\overline{W}$ (write 0). At the cell terminal to which $W(\overline{W})$ is applied, there is obtained the read $R(\overline{R})$ or sense $S(\overline{S})$ data output. Such a memory unit is indicated schematically in Fig. 9-10.

The basic elements of which a RAM is constructed are indicated in Fig. 9-9. These include the rectangular array of storage cells, the $X$ and $Y$ decoders, the write amplifiers to drive the memory, and the sense amplifiers to detect (read) the stored digital information. The amplifiers labeled $R/W$ 0 and $R/W$ 1 are not indicated explicitly in Fig. 9-9 (but are drawn in Fig. 9-12).

The organization (also called a *functional diagram*) for a 4096 word by 1-bit read/write memory is given in Fig. 9-11. Note that there are 64 rows and 64 columns in the memory array. Hence, each decoder has 6 inputs. The data

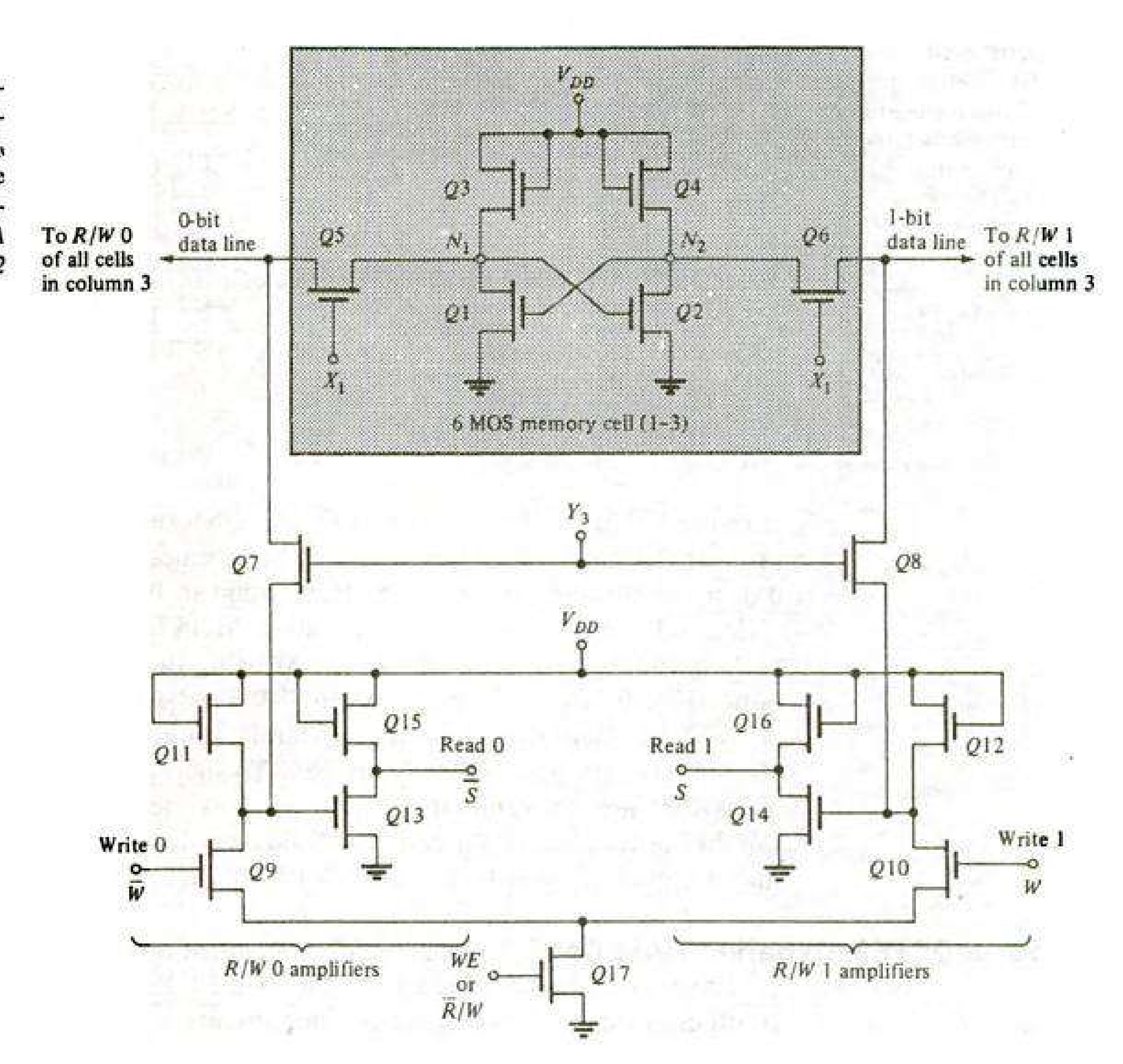



**FIGURE 9-12**
A storage cell (1–3) containing 6-NMOS transistors. The $X_1$ and $Y_3$ address lines and the write and read amplifiers are also shown. *A logic 1 is stored if Q2 conducts.*

and the 0-bit data line is at $V_{DD}$, $Q13$ is ON, and $\bar{S} = 0$ V. We thus have correctly sensed that the 1-3 FLIP-FLOP stores a 1 (since $S = 1$ and $\bar{S} = 0$).

*In order to write a 1 into the cell* we address it ($X_1 = 1$ and $Y_3 = 1$), set $WE = 1$, $W = 1$, and $\overline{W} = 0$. Then $Q17$ and $Q10$ are ON and $Q9$ is OFF. Hence, the 1-bit line is grounded and the 0-bit line goes to $V_{DD}$ through load $Q11$. Current now flows into the 1-bit line from $V_{DD}$ through $Q4$, $Q6$, $Q8$, $Q10$, and $Q17$ to ground. Thus node $N_2$ is effectively grounded. This cuts off $Q1$, and $N_1$ rises to $V_{DD}$. Consequently $Q2$ is held ON and $N_2$ is maintained at 0. When the address is removed ($Q5$, $Q6$, $Q7$, and $Q8$ are OFF), $Q2$ is ON, $Q1$ is OFF, and a 1 has been written into the selected memory cell.

**Static CMOS RAM Cell** The static CMOS RAM cell is similar in structure and operation to the NMOS cell depicted in Fig. 9-12. The CMOS circuit in Fig. 9-13 corresponds to the 6-MOS memory cell in Fig. 9-12 with identical transistor numbering. The sense amplifiers required to read and write data are not shown. Transistors $Q1$ to $Q4$, in Fig. 9-13, are the cross-coupled CMOS inverters which form the

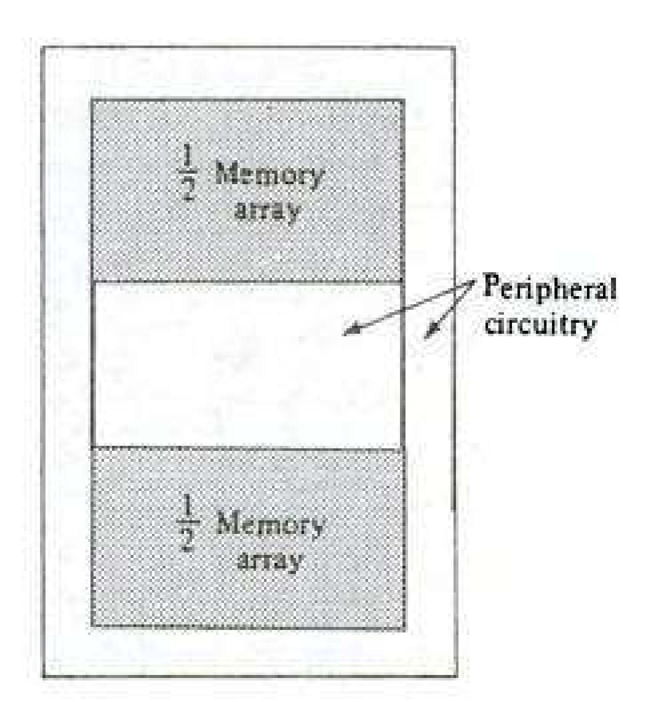



**FIGURE 9-16**
Typical chip layout of a 64-kb dynamic RAM (DRAM).

**Dynamic RAM Chip Organization** Typical 16- to 64-kb DRAMs have chip layouts similar to that shown in Fig. 9-16. The memory array is divided into two equal segments and peripheral circuitry is placed along the borders of each array segment. The memory-support circuits include the row and column decoders and buffers, the sense and refresh amplifiers, input and output registers, and timing and control circuits. Dynamic RAMs generally come in the industry standard 16-pin DIP package.[1] For a 64-kb DRAM, such as the Mostek 4164 or Texas Instruments TMS 4164 (these are pin-compatible), the eight row and eight column address lines are multiplexed on eight pin connections. This is accomplished by adding two externally generated clock signals called the *row-address strobe* (RAS) and the *column-address strobe* (CAS). Actually, the complements of these signals, $\overline{RAS}$ and $\overline{CAS}$, are applied to the pins and latches the row and column addresses onto the chip. In addition to the eight address lines, 7 of the 8 pins remaining are used for the two strobe signals, the write enable signal ($\overline{WE}$), the data input and output lines, ground, and the 5-V supply. No connection is made to the last pin.

Larger-capacity DRAMs, such as the Texas Instruments 256-kb × 1-bit TMS4256, utilize chips organized as displayed in Fig. 9-17. This arrangement is easily identified in the photomicrograph of Fig. 9-18. The memory array is partitioned into four 64-kb arrays, each of which is organized as shown in Fig. 9-16. The nine row- and nine column-address lines needed to select 1 of 262,144 bits are latched onto the chip by means of the row-address and column-address strobes ($\overline{RAS}$ and $\overline{CAS}$). The 16 pin connections are the nine address lines, the two data lines, the two strobe lines, the write-enable signal ($\overline{WE}$), ground, and the 5-V supply. The timing and control signals and data lines are all TTL-compatible.

The basic organization in Fig. 9-17 can be modified so that the 256-kb capacity provides storage of 64K 4-bit words (64K × 4). Such a chip is shown (the Texas Instruments TMS 4464) in Fig. 9-19 and is available in an 18-pin DIP

[1]Many chips also are available in an 18-pin plastic chip-carrier package.

conducting BJT from the row line ($X$) to the appropriate data line ($D_0$ or $D_1$).

To write a 1, $X$ and $X^*$ are set to $V(1)$ and $D_1 = V(0)$. Thus the emitter $E_1$ of $Q1$ is forward-biased and current exists in $Q1$. The voltage $V_{C1} = V_{B2}$ decreases, and, with $D_0 = 1.5$ V and $X = V(1)$, both emitter junctions of $Q2$ are reverse-biased and $Q2$ is OFF. When voltage levels return to their standby levels, $X = V(0)$ and $D_0 = D_1 = 1.5$ V, $Q1$ remains ON as $E_{X1}$ is forward-biased. Sufficient base current for $Q1$ is supplied through $R_2$. Although the emitter $E_{X2}$ of $Q2$ is low, $X^* = V(0)$ causes $V_{C1} = V_{B2}$ to decrease from its value when $X^* = V(1)$, and the lower value of $V_{B2}$ virtually eliminates any base current in $Q2$. Thus it is reasonable to assume that $Q2$ is OFF. A 1 is stored in the cell because $Q1$ is ON and a current path exists in the $X$ line (through $E_{X1}$).

With $D_1 = V(1)$, reading the stored 1 is accomplished by making $X$ and $X^* = V(1)$. These voltage values reverse-bias the emitters $E_{X1}$ and $E_{X2}$. Because $X^* = V(1)$, sufficient base drive through $R_2$ is established to forward-bias the $E_1$ junction so that $Q1$ still conducts. The current path, however, is switched to the $D_1$ line and returns to ground through $R_3$ and the 1.5-V supply. The voltage at $D_1$ increases as a result of the voltage drop in $R_3$, and sensing this higher voltage indicates the presence of a 1.

To write a 0, $X = V(1)$ and $D_0 = V(0)$. The conditions that exist on $E$ and $E_0$ are reversed from those encountered for writing a 1. The operation of the circuit is the same except that it is $Q2$ that conducts with the current path through $E_{X2}$ during standby. The stored 0 can be read from memory by making $X$ and $X^* = V(1)$ with $D_0$ and $D_1$ remaining at 1.5 V. By analogy, the 0 is sensed because of the increased voltage drop in $R_4$ caused by the current switching to the $D_0$ line.

The circuit in Fig. 9-21 is a second BJT memory cell and is generally used when the bipolar process utilized permits fabrication of the Schottky diodes

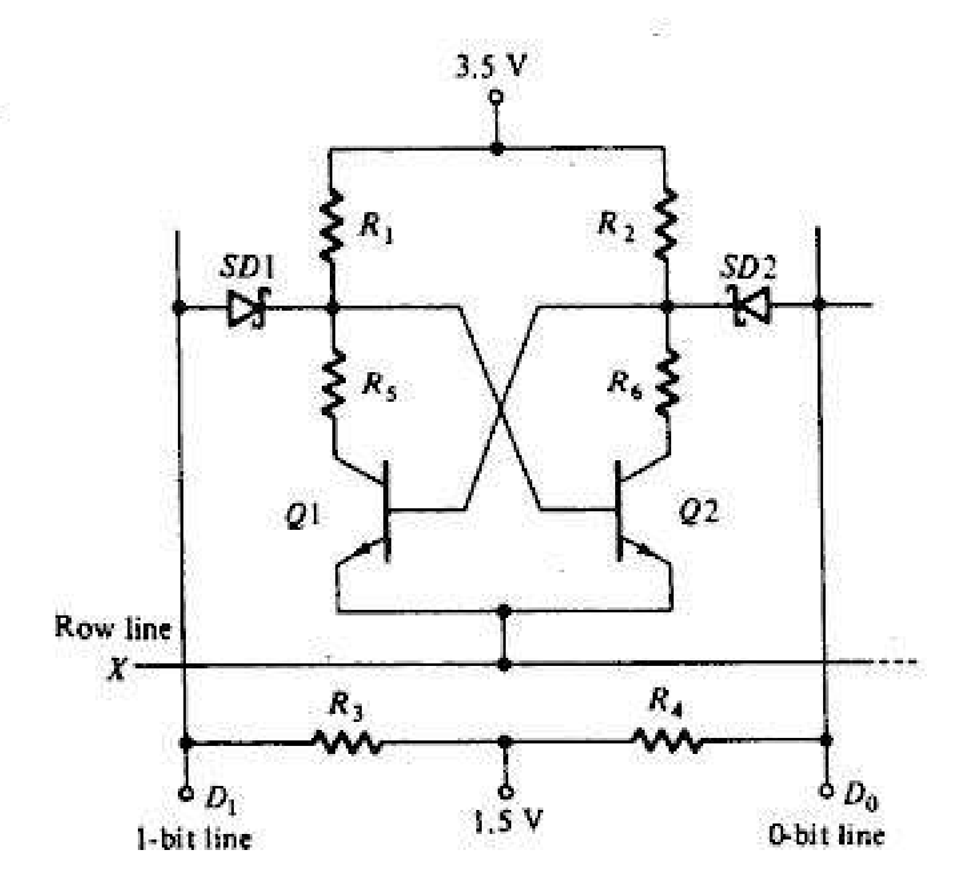



**FIGURE 9-21** A diode-coupled bipolar memory cell.

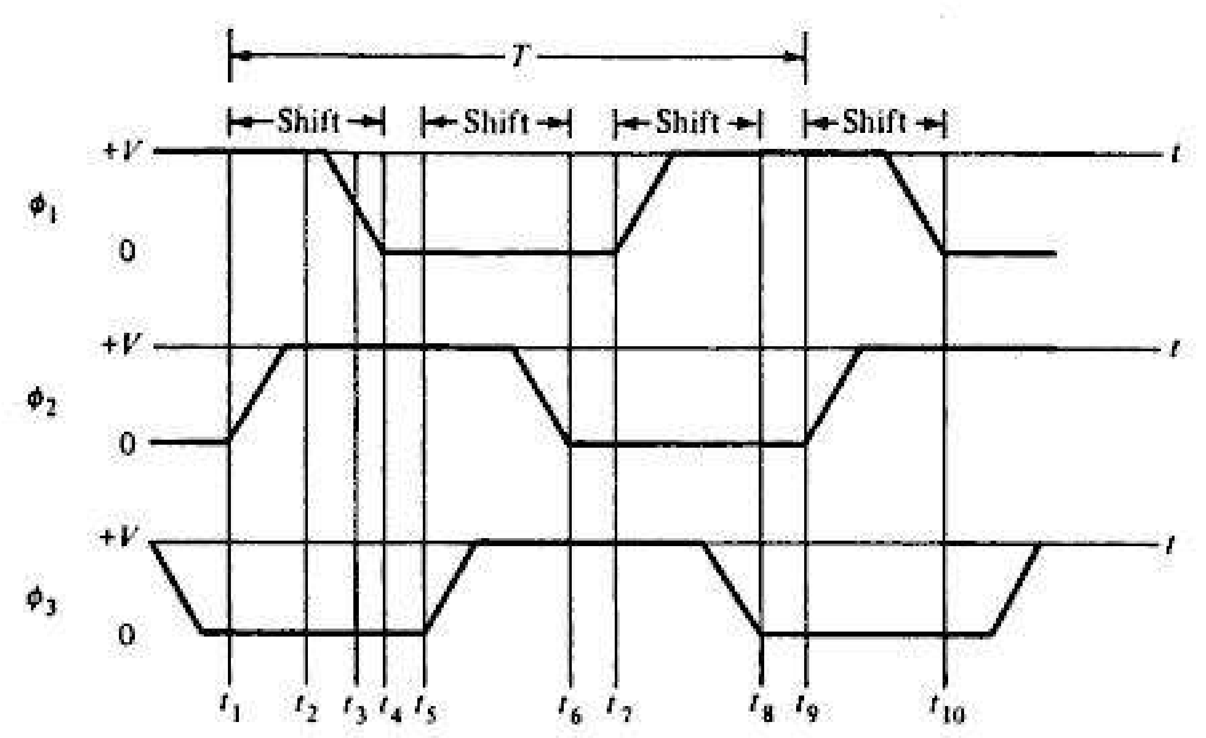



**FIGURE 9-24**
The three-phase excitation waveforms for the CCD in Fig. 9-23*a*. The potential profile in Fig. 9-23*b* corresponds to the instant $t_1$ of this figure; Fig. 9-23*c* corresponds to $t_2$; and so on.

9-24 should be at a voltage in excess of $V_T$ for the electric field to penetrate into the channel and form the depletion region.

**Minimum and Maximum Operating Frequencies** Steady-state (dc) operation of a CCD is not possible. Thermally generated carriers become trapped in empty potential-energy wells and, in time, change the logic state from a 0 to a 1. This phenomenon, called the *dark-current effect*, sets a lower limit to the clock frequency (10 kHz to 1 MHz).

No steady-state power is required by a CCD cell since power is dissipated only in charging the effective cell capacitances. Consequently the upper limit of clocking frequency (1 to 30 MHz) may be determined by the maximum allowable power dissipation. Also, an increase in frequency reduces the efficiency of transfer of charge from one cell to the next. Hence, the upper frequency may be limited by the point at which transfer losses become unacceptable.

## 9-8 CCD STRUCTURES

A CCD cannot be assembled from discrete components because a single continuous channel is required to provide the coupling between depletion regions. The gates (Fig. 9-23) must be separated by a small distance (~1 μm) to supply this coupling, and these narrow gaps are difficult to fabricate reliably. A number of alternative structures with both metallic and polysilicon gates have been developed to avoid difficulties in fabrication. One such polysilicon electrode structure is depicted in Fig. 9-25 for a three-phase, *n*-channel CCD. This planar electrode structure makes use of overlapping gates having different shapes. A large number of such CCDs are fabricated in rows parallel to one another to cover the chip area. With minimum separation between rows (2 to

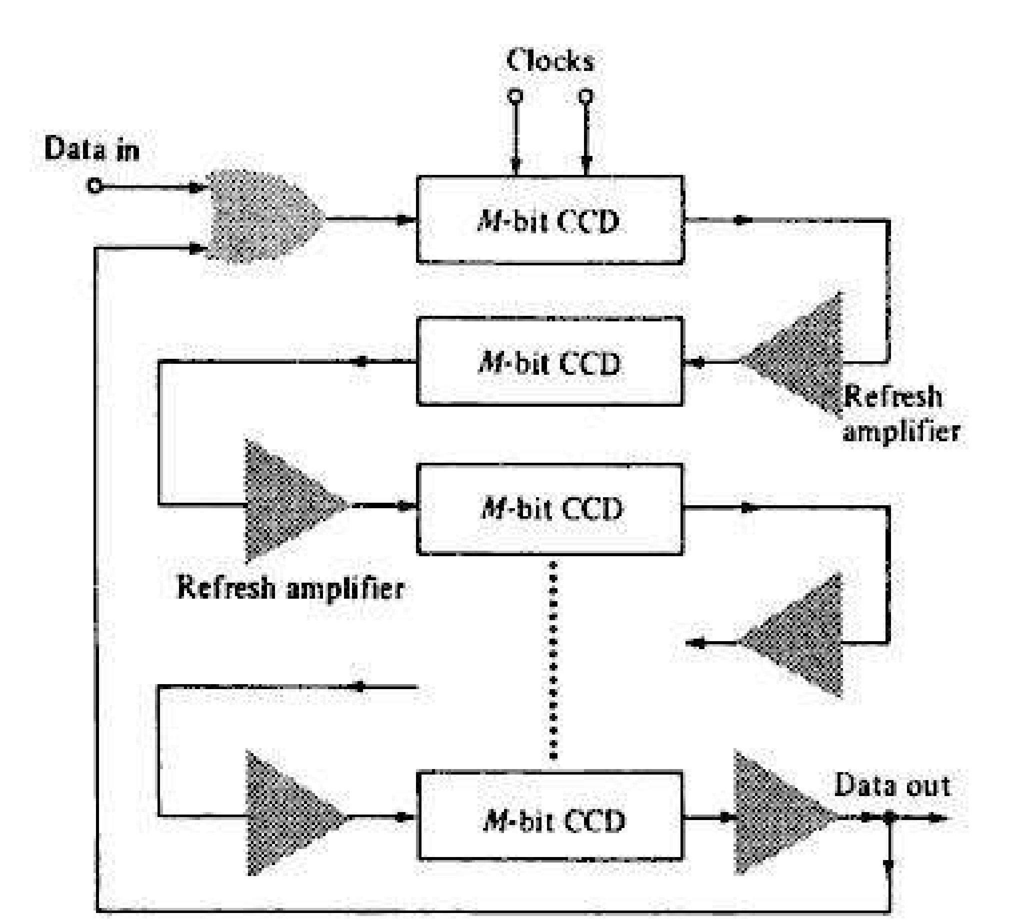



**FIGURE 9-30**
Serpentine or synchronous organization of a CCD memory. The multiphase clock is applied simultaneously to all CCD sections (between refresh amplifiers).

The serpentine is the simpler of the two CCD memories to fabricate and is illustrated in Fig. 9-30. It is a synchronous organization in which data is shifted from cell to cell in a long, snakelike manner (hence the name "serpentine") in a recirculating shift register.

The LARAM is designed for short access times and consists of a number of short recirculating memories operating in parallel. These parallel CCDs share common input and output lines. A decoder is used to address the registers at random and gives rise to the designation of this organization. The Intel 2464 is a 64-kb CCD organized as 256 independent registers of 256 bits each (Fig. 9-31). An 8-to-256-line decoder can select any register at random. The I/O operations are performed in a manner similar to that for a 256-bit RAM.

## 9-9 INTEGRATED-INJECTION LOGIC ($I^2L$)

*Integrated-injection logic* or simply $I^2L$, developed simultaneously in 1972 by engineers at Phillips Research in the Netherlands and IBM Laboratories in West Germany, is a VLSI technology that combines the higher component density of MOS with the higher speeds of BJTs. Advances in MOS technology, such as the reduction of minimum dimensions, have lead to the declining use of $I^2L$. However, $I^2L$ RAMs, microprocessors, and A/D and D/A converters are currently available. In addition, research and development in this technology is proceeding, albeit at a diminished level compared to the decade since its introduction. For these reasons we include this section to introduce the principles underlying $I^2L$ operation.

**Merging of Devices** Increased component density in bipolar fabrication can come about by eliminating area-consuming resistors and significantly reducing (or eliminating) the isolation islands that separate devices. One such technique, already en-

inversion has been performed by $Q1$ because $V_o = 0.75$ V for $V_i = 0$ and $V_o = 0$ V for $V_i = 0.75$ V. The logic swing is about 0.75 V, its exact value depending on the bias current $I_j$.

Note that in saturation the collector current is $I_j$ and the base current has about the same magnitude. Hence, a value of the $CE$ current gain $\beta_F$ of only unity is required to cause saturation. A transistor operating in the inverse mode has a very much smaller value of $\beta_F$ than in the normal mode (~ 100). Nevertheless, a $CE$ current gain in excess of unity (approximately 2 to 10) is easily obtained for the upside-down transistor.

When an inverter is switched from one state to the other, the voltages of the transistor capacitances must change values, causing a propagation delay $t_{pd}$. The charging (and discharging) current of these capacitances is supplied by the injector. Large values of $I_j$ result in small values of $t_{pd}$, but the penalty which must be paid for the decreased $t_{pd}$ is an increased average-power dissipation $P_{av}$.

An advantage of the $I^2L$ configuration is that it can be operated over a wide range of speeds by simply altering the total injector current, by varying the one resistor $R_x$. The range of operation extends from about 1 nA to 1 mA. Thus after a chip is designed and built, the desired speed of operation may be selected by changing $R_x$.

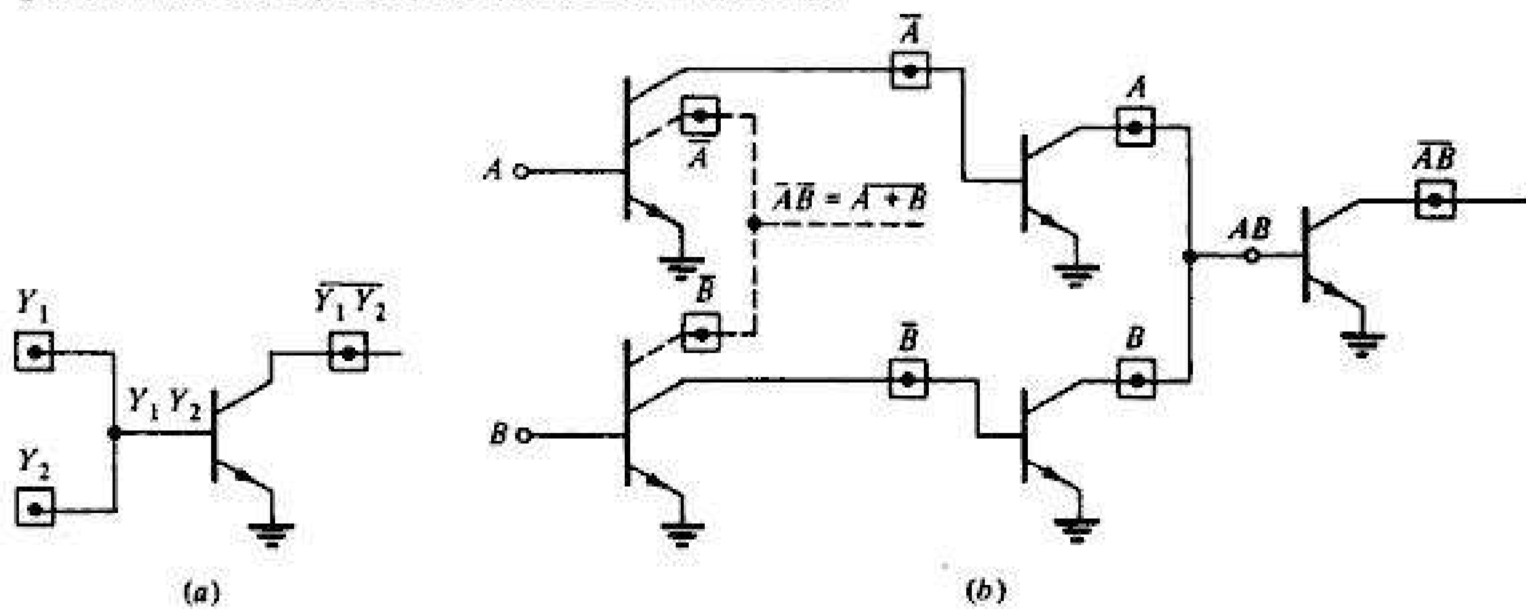
**NAND Gate**

To obtain an AND gate in injector logic is extremely simple. In Fig. 9-36*a*, $Y_1$ is a logic variable at the output of an $I^2L$ inverter and $Y_2$ is another variable at the collector of a second $I^2L$ gate. Connecting $Y_1$ and $Y_2$ together yields $Y = Y_1Y_2$ at the common node in Fig. 9-36*a*. If $Y$ is applied to the input of an inverter, the output is the NAND function $Y_1Y_2$, as shown in Fig. 9-36*a*.

If $A(B)$ is an externally applied logic variable, then to obtain $A(B)$ *at the collector of an $I^2L$ gate*, two cascaded inverters must be used, as indicated by the solid lines in Fig. 9-36*b*. Proceeding in this manner, the NAND function $\overline{AB}$ is obtained as shown in the figure.

**FIGURE 9-36**

(*a*) A NAND gate using wired AND for internal (collector) logic variables. (*b*) A NAND gate for externally applied logic variables (solid connections). The dashed portion of the circuit is a NOR gate. (*Note:* The injectors are omitted for simplicity.)



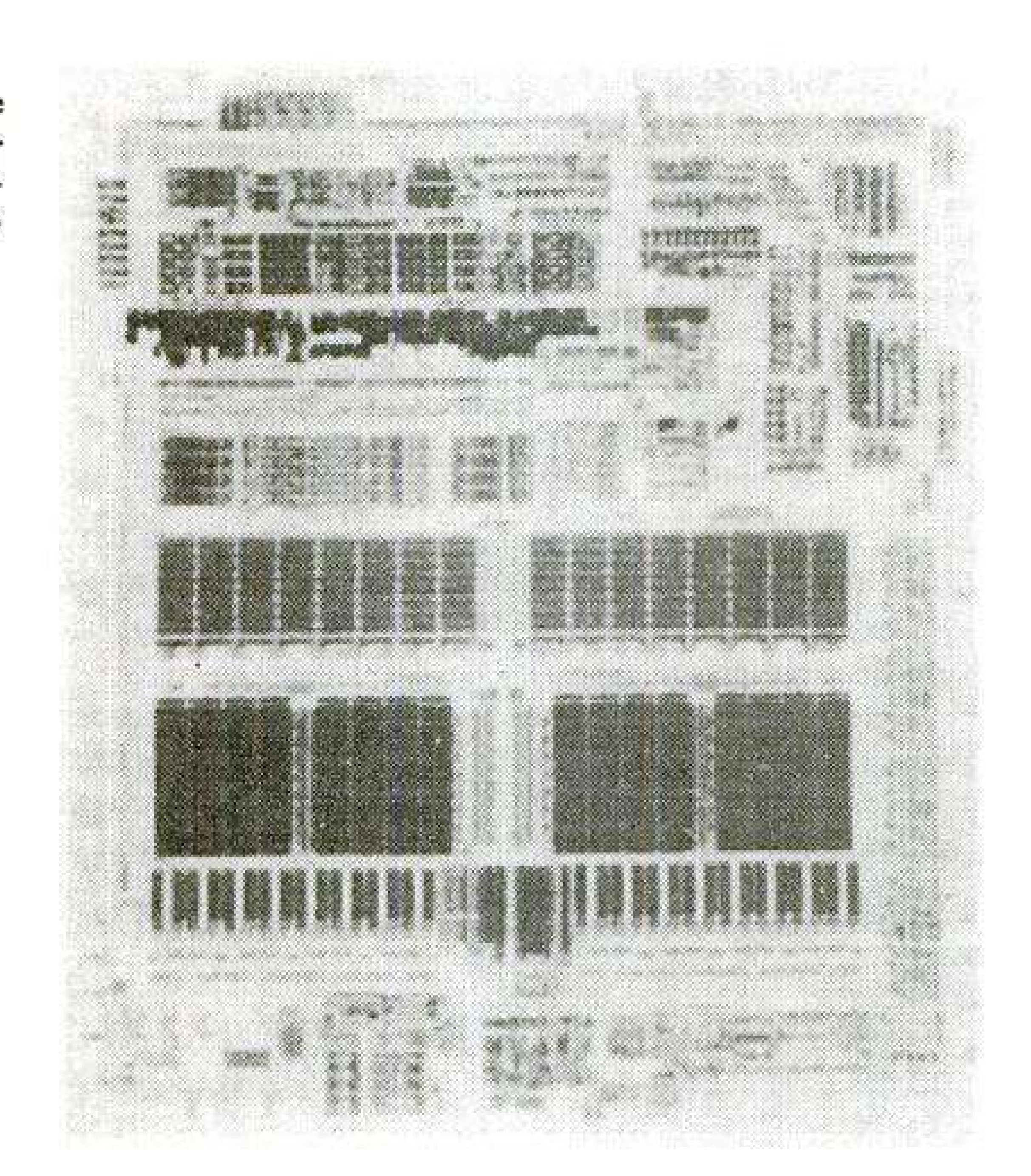


**FIGURE 9-41**
Photomicrograph of the MC68HC11, a single-chip microcomputer. (*Courtesy of Motorola, Inc.*)

is displayed in Fig. 9-41, is one such chip. Containing over 100,000 transistors on a 6.5 × 7.4 mm (256 × 287 mil) die, this microcomputer comprises a CPU, timer, I/O, memory (RAM, ROM, EEPROM), and an A/D converter. Figure 9-41 also illustrates the relative size required to implement the three types of memory. Starting from the bottom of the chip, the first step contains 512 bytes of EEPROM. The 8192 bytes of ROM form the second strip, while the third strip is for the 256 bytes of RAM. The single-chip microcomputer can be augmented by other chips to both expand memory capacity and provide the interfaces with a variety of I/O and communications devices.

**Applications**

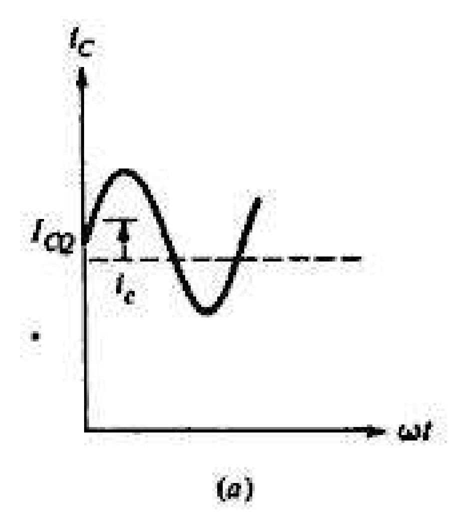
A detailed discussion of microprocessor applications is beyond the scope of this book. Indeed, today most electrical and computer engineering departments offer (and often require) courses in digital design, computer architecture, and microprocessor-based design where these applications are treated. However, the list on the following page indicates the range of applications which incorporate microprocessors.

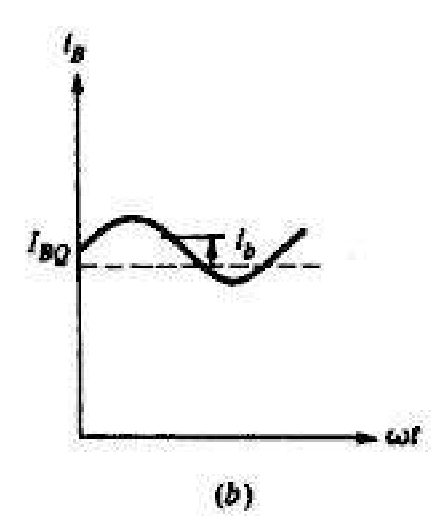
Digital-signal processing is another area in which microprocessors are incorporated. Single-chip programmable signal processors, such as the Texas Instruments TMS 32010, are capable of performing the real-time functions of

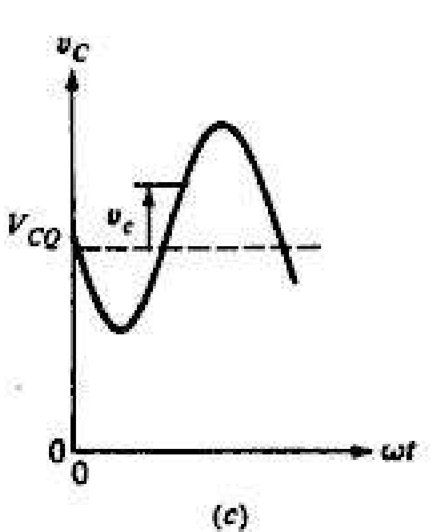
# *Part Three*
# AMPLIFIER CIRCUITS AND SYSTEMS

The availability of reliable devices to amplify signals is the foundation on which modern electronics rests. Without amplifying devices, nearly all of today's communication, control, instrumentation, and computer systems would be rendered impractical. In Secs. 3-1 and 4-1 we demonstrated that the controlled source is an amplifying element. Also, in those chapters it was seen that both BJTs and FETs, when appropriately biased, behave as controlled sources. In this part of the text we examine how transistors are utilized in amplifier stages and in turn, how such stages are interconnected in amplifier systems. When providing controlled-source characteristics, transistors behave linearly. Consequently, the analyses of amplifier circuits lend themselves to the use of small-signal, incremental models (Secs. 3-10 and 4-14).

The first two chapters in this part deal with the behavior of basic transistor amplifier configurations. In Chap. 10 we focus on low-frequency circuit performance and the biasing process. The frequency response, and the resultant limitations on performance of these basic circuits, is treated in Chap. 11. The very important topic of feedback and its use in controlling circuit performance is discussed in Chaps. 12 and 13. The feedback concept and low-frequency feedback amplifier behavior is treated in Chap. 12; stability and frequency response are discussed in Chap. 13. The objective in Chap. 14, where we describe in detail the operational amplifier, the predominant analog integrated circuit (IC), is twofold: (1) this single-chip amplifier system is an essential component in signal-processing and data-acquisition circuits; thus this treatment serves as a prelude to Part 4; and (2) amplifier design techniques are readily highlighted and the material in the four prior chapters is integrated by the discussion of the operational amplifier.



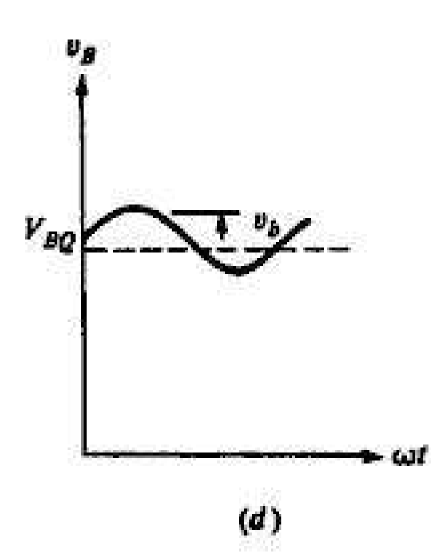




**FIGURE 10-3**
Sinusoidal waveforms for transistor currents and voltages: (*a*) collector current $i_C$, (*b*) base current $i_B$, (*c*) collector voltage $v_C$, (*d*) base voltage $v_B$. [*Note*: Each quantity comprises a constant (dc) term and a sinusoidally varying (ac) term.]

Electrical and Electronic Engineers (IEEE). These are given in Table 10-1 for the bipolar transistor and FET symbols are listed in Table 10-2.

The system underlying the notation is as follows:

1. Lowercase letters $i$ and $v$ designate instantaneous ("Inst." in Tables 10-1 and 10-2) current and voltage, respectively.
2. Uppercase $V$ and $I$ denote rms (or peak) values of ac components or average values (dc) of total quantities.
3. Lowercase subscripts are used only with the time-varying components of voltage and current.
4. Uppercase subscripts refer to total or dc quantities.
5. Repeated subscripts refer to the magnitudes of supply quantities.
6. Currents are taken as positive when entering a terminal from the external circuit.
7. Voltages are taken as positive when measured with respect to the reference node (usually ground). For voltages measured between a pair of terminals (not the reference), an additional subscript is used. Thus $v_{CE}$ is the instantaneous collector-to-emitter potential.
8. The additional subscript $Q$ is added to circuit variables to indicate quiescent values.

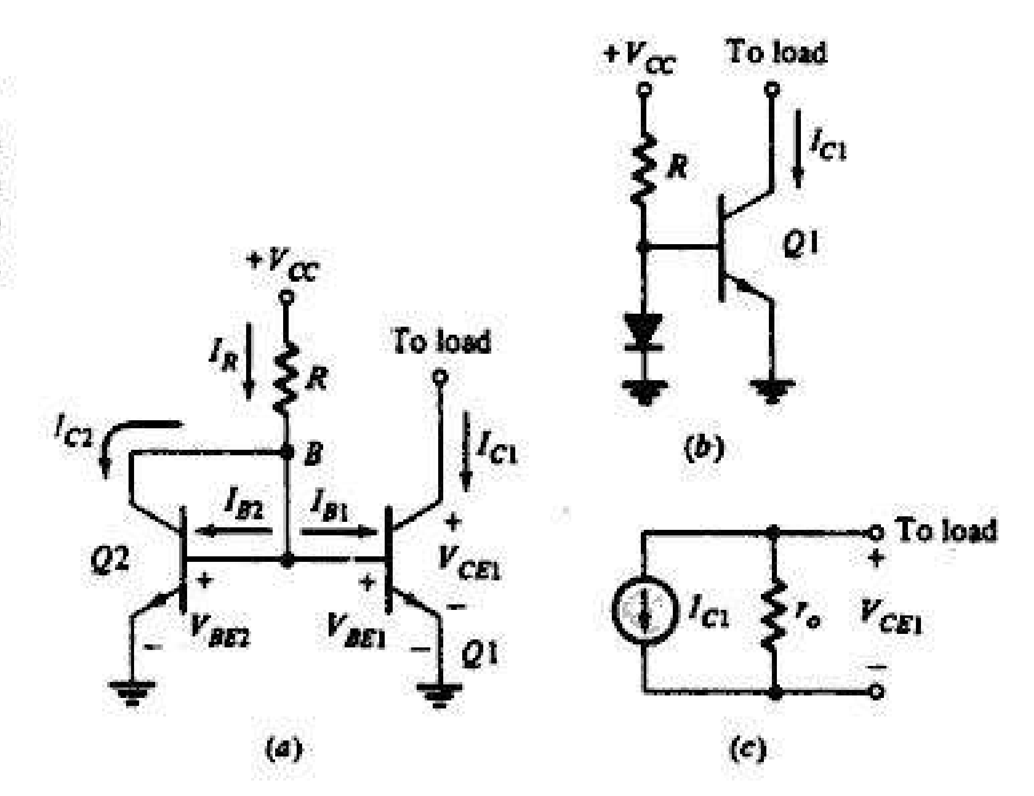



**FIGURE 10-5**
(*a*) Current mirror circuit. (*b*) Representation of (*a*) with *Q*2 replaced by equivalent diode. (*c*) Norton equivalent of the current mirror.

imum and maximum values of $\beta_F$ for a given transistor over the operating temperature range. For the circuit in Fig. 10-1*a*, the operating point can lie anywhere between $Q_1$ and $Q_2$ (Fig. 10-1*b*) for $50 \le \beta_F \le 125$. Clearly, on the basis of the output waveform in Fig. 10-4*b*, this bias arrangement is inappropriate. Proper circuit performance requires that the location of the operating point be controlled; that is, biasing conditions must be stabilized. To achieve bias stability, we must restrict the operation to a small region of the load line over the specified range of $\beta_F$ so that the signal (say, $i_b = 20 \sin \omega t\ \mu A$) is accommodated. The import of the previous sentence is that *effective bias control occurs when the collector current, and consequently the emitter current, remains essentially constant and independent of variations in* $\beta_F$. In Fig. 10-1*b* we see that maintaining the collector current at or slightly larger than 3 mA restricts the operation of the BJT to the forward-active region (Figs. 10-2*b* and 10-4*b*) for both $\beta_F = 50$ and $\beta_F = 125$. The following five sections describe circuits and techniques used to obtain bias stability.

**10-3 BJT BIASING FOR INTEGRATED CIRCUITS** The circuit in Fig. 10-5*a* is typical of the biasing arrangement used in ICs. This and similar circuits are called *current sources* or *current mirrors* and are designed to maintain the current $I_{C1}$ at a constant value.[1] Recall that we encountered this circuit in Sec. 3-11 and used it in the differential amplifier stage (emitter-coupled stage) of Sec. 3-12 and again in conjunction with ECL gates (Sec. 6-14). The differential amplifier is one of the principal building blocks in IC amplifiers and is treated in detail in Sec. 10-15. In Fig. 10-5*a* we recognize that *Q*2 is connected as a diode as shown in Fig. 10-5*b*. The circuit in Fig. 10-5*c* is the equivalent representation of the current source. The resistance $r_o$ is the output resistance of the source and reflects the fact that practical sources are nonideal.

[1]This circuit was investigated in Example 3-8.

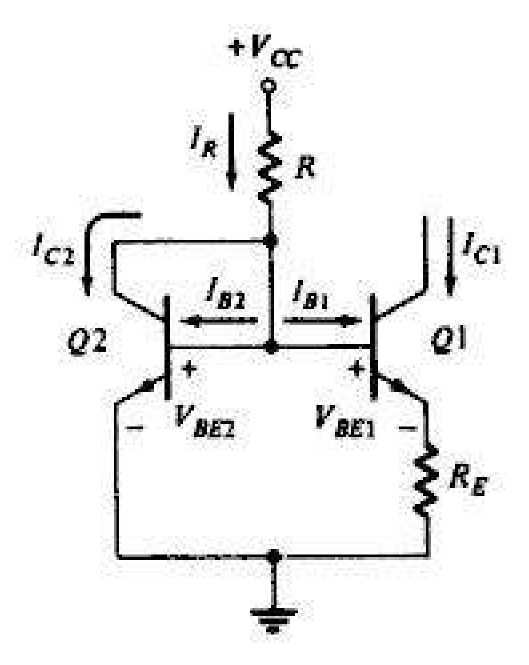



**FIGURE 10-8**
The Widlar current source.

In this configuration $V_{BE1}$ is less than $V_{BE2}$, and, consequently, $I_{C1}$ is smaller than $I_{C2}$ [Eq. (10-4)]. The asymmetric nature of the base-emitter loop causes this circuit to act as a "lens" rather than a "mirror." In effect, $Q2$, $V_{CC}$, and $R$ establish the reference current $I_R$, and the value of $R_E$ determines the degree to which $I_{C1}$ is less than $I_R$.

The KVL equation for the emitter-base loop in Fig. 10-8 is

$$V_{BE2} = V_{BE1} + (I_{B1} + I_{C1})R_E$$

or

$$V_{BE2} - V_{BE1} = \Delta V_{BE} = (I_{B1} + I_{C1})R_E \tag{10-10}$$

As indicated in Eq. (10-3), the collector current of a transistor depends heavily on the base-emitter voltage. For identical *npn* transistors, $I_{C1}$ and $I_{C2}$ are given by Eqs. (10-3) and their ratio in Eq. (10-4). Equation (10-4) can be rewritten by taking the natural logarithm of both sides as

$$V_{BE2} - V_{BE1} = \Delta V_{BE} = V_T \ln \frac{I_{C2}}{I_{C1}} \tag{10-11}$$

Equating $\Delta V_{BE}$ in Eqs. (10-10) and (10-11) results in

$$R_E = \frac{V_T}{I_{C1}\left(1 + \dfrac{1}{\beta_F}\right)} \ln \frac{I_{C2}}{I_{C1}} \tag{10-12}$$

The reference current $I_R$ as given in Eq. (10-6) as the KVL expression for the loop containing $V_{CC}$, $Q2$, and $R$ is the same in both Figs. 10-5*a* and 10-8. The KCL expression

$$I_R = I_{C2} + I_{B2} + I_{B1}$$

is rewritten as

$$I_R = I_{C2}\left(1 + \frac{1}{\beta_F}\right) + \frac{I_{C1}}{\beta_F} \tag{10-13}$$

output resistance above that obtained in the Wilson source while ensuring that the source current is independent of $\beta_F$ variations.

**10-6 DISCRETE-COMPONENT BJT BIASING—ANALYSIS** Bias stabilization is as important a consideration in transistor stages constructed from discrete components as it is in ICs. However, as matched pairs of BJTs are costly and since no restrictions exist on resistance values, current sources are not used. The design objective still remains, keeping the collector current constant as $\beta_F$ varies. To achieve bias stability, base current is allowed to vary with $\beta_F$. In Fig. 10-1*b*, to maintain $I_C \approx 3$ mA independent of $\beta_F$ variations, requires that $I_B$ decrease as $\beta_F$ increases. The four-resistor circuit of Fig. 10-12 is the most widely used discrete-component biasing arrangement.

**Capacitive Coupling** In Fig. 10-12, the capacitors $C_{B1}$ and $C_{B2}$ are called *blocking* or *coupling capacitors*. The capacitance $C_{B1}$ is used to couple the signal from the input source $v_s$ to the transistor, and $C_{B2}$ couples the output signal from the BJT to the load $R_L$. Under quiescent conditions $C_{B1}$ and $C_{B2}$ act as open circuits because the reactance of a capacitor is infinite at zero frequency (dc). The values of these capacitors are chosen sufficiently large that, at the lowest signal frequency, their reactances are small enough that they may be considered short circuits. This serves to isolate $v_s$ and $R_L$ from the four resistances $R_1$, $R_2$, $R_C$, and $R_E$ used to establish the bias. These capacitors block dc and freely pass signal voltages. For example, the quiescent collector voltage does not appear at the output, but $v_o$ is an amplified replica of the input signal $v_s$. The output signal voltage often serves as the input to another amplifier stage ($R_L$ is the input resistance of that stage) without affecting its bias because of the blocking effect of $C_{B2}$.

The capacitance $C_E$ (called a *bypass* capacitor) is also selected so that it may be treated as a short circuit at the lowest signal frequency. Thus, for

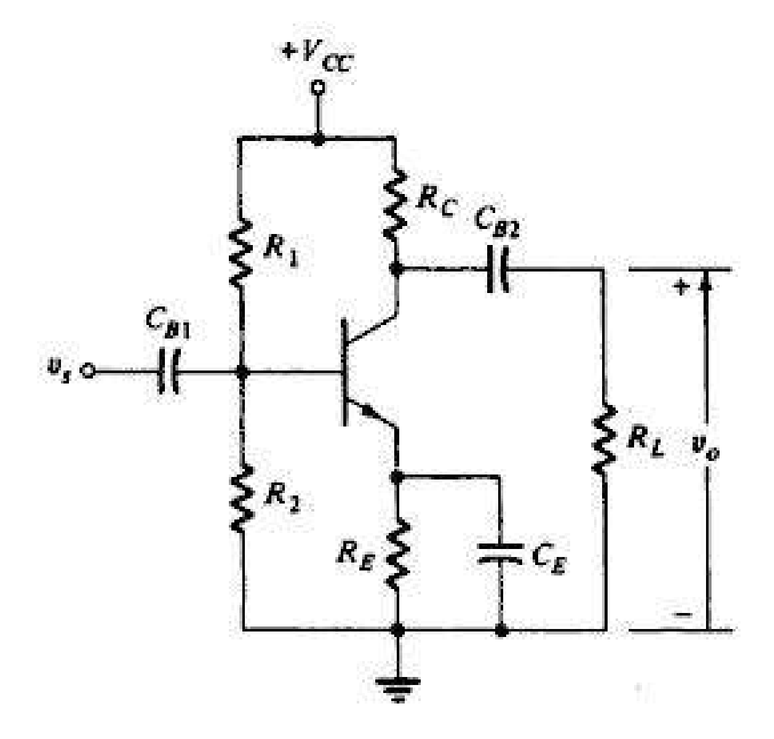



**FIGURE 10-12**
A discrete-component common-emitter amplifier stage.

**Current Increment for a Change in $V_{BE}$** From Eq. (10-22) with $\beta_F \gg 1$, it follows, if $\beta_F$ and $I_{CO}$ remain constant, that

$$\Delta I_C = -\frac{\beta_F}{R_B + \beta_F R_E}\Delta V_{BE} = -\frac{M_1}{R_E}\Delta V_{BE} \tag{10-27}$$

Note that in Eqs. (10-26) and (10-27) it is assumed that a BJT having $\beta_{F1}$ is used.

**Total Current Increment** To obtain the total change in current over a specified temperature range due to simultaneous variations in $\beta_F$, $I_{CO}$, and $V_{BE}$, we add the individual increments found in Eqs. (10-25), (10-26), and (10-27). The fractional change in collector current is given by

$$\frac{\Delta I_C}{I_{C1}} = \left(1 + \frac{R_B}{R_E}\right)\frac{M_1 \Delta I_{CO}}{I_{C1}} - \frac{M_1 \Delta V_{BE}}{I_{C1} R_E} + \left(1 + \frac{R_B}{R_E}\right)\frac{M_2 \Delta\beta}{\beta_{F1}\beta_{F2}} \tag{10-28}$$

where $M_1(M_2)$ corresponds to $\beta_{F1}$ ($\beta_{F2}$). Note that as $T$ increases, $\Delta I_{CO}/I_{C1}$ and $\Delta\beta$ increase, whereas $\Delta V_{BE}/I_{C1}$ decreases. Hence all terms in Eq. (10-28) are positive for increasing $T$ and negative for decreasing $T$.

## 10-7 DISCRETE-COMPONENT BIASING—DESIGN

The following example illustrates the design technique for discrete-component biasing. It is used also to examine the relative sizes of the three components of $\Delta I_C/I_{C1}$ in Eq. (10-28). The transistor values used are typical of those encountered in practice. Many circuits for commercial applications are required to operate from 0 to 70°C; others, such as those used in aerospace, automotive, and military applications, often must operate from −55°C to between 100 and 150°C. Manufacturers frequently provide device data for the range −65 to 175°C to accommodate the wide variety of applications.

*Example 10-4* A 12-V supply is used to bias the transistor in the circuit shown in Fig. 10-12. On the basis of the signal to be amplified, it is determined that $1.00 \le I_C \le 1.15$ mA and $5.0 \le V_{CE} \le 6.0$ V. The collector resistance $R_C = 1.5$ kΩ and is selected to achieve the desired gain. The BJT parameters are

| | | |
|---|---|---|
| $T = -55°\text{C}$ | Lowest value of $\beta_F = 40$ | $V_{BE} = 0.88$ V |
| $T = +125°\text{C}$ | Highest value of $\beta_F = 400$ | $V_{BE} = 0.48$ V |

(*a*) Design the circuit ($R_1$, $R_2$, and $R_E$) to meet the specifications. Consider that the effect of $I_{CO}$ variations is negligible.

(*b*) Using values obtained in part *a*, determine the component of the total current increment $\Delta I_C$ attributed to changes in $I_{CO}$ with temperature. Manufacturers' specifications indicate $I_{CO} = 2.0$ pA at $T = -55°\text{C}$ and $I_{CO} = 525$ nA at $T = 125°\text{C}$.

load line caused by the input signal $\Delta i_B$ about $I_{BQ}$ determines the output signal $\Delta v_{CE}$. If $V_{CC}$ and the $Q$ point are specified, the static load line is determined uniquely. However, a reduction in $R_C$ and consequently $R'_L$ (so $R_E$ may be increased) causes the slope of the dynamic load line to increase (it becomes more vertical). For a given $\Delta i_B$, the projection of this line segment on the $V_{CE}$ axis decreases ($\Delta v_{CE}$ decreases). The reduced output signal for a given input signal is indicative of a decreased gain (amplification) in the stage.

## 10-8 FET BIASING

Biasing techniques which stabilize the $Q$ point in IC and discrete-component FET circuits parallel those discussed in the previous two sections for the BJT. In MOS circuits, biasing schemes control deviations in the operating point caused by fabrication variations in the threshold voltage $V_T$ and transconductance (processing) parameter $k$. Both IC and discrete-component JFET circuits are biased so that the effects of unit-to-unit variations in the pinch-off voltage $V_P$ and zero-bias drain saturation current $I_{DSS}$ are controlled. Both MOSFETs and JFETs are operated in their saturation regions at all times so that they provide controlled-source characteristics.

### Current Sources

The circuit in Fig. 10-15*a* is a MOS IC current mirror. The transistor $Q1$ supplies the load with a current $I_{D1}$. The reference current $I_R = I_{D2}$ is provided by $V_{DD}$, $R$, and the enhancement-load transistor $Q2$ (Sec. 4-11). These components also determine the value of $V_{DS2} = V_{GS2}$, and because the gates are tied together, $V_{GS1} = V_{GS2}$. The values of $V_T$ and $k$ are identical for $Q1$ and $Q2$ since they are fabricated simultaneously. Hence, for MOSFETs having the same aspect ratio ($W/L$), the currents $I_{D1}$ and $I_{D2}$ are equal.

***Example 10-5*** Two identical transistors having characteristics given in Fig. 4-12 and repeated in Fig. 10-15*b* are used in the circuit in Fig. 10-15*a*. The supply voltage is 6 V, and $R = 20$ kΩ. Determine the source current.

***Solution*** The enhancement-load characteristic (shown in black) is drawn in Fig. 10-15*b* and is the locus of points for which $V_{DS2} = V_{GS2}$. The load line corresponding to $V_{DD} = 6$ V and $R = 20$ kΩ is also displayed in Fig. 10-15*b*. The value of $I_{D2} = I_R = 90\ \mu$A is obtained at the intersection of the load line and the resistance characteristic. Because $Q1$ and $Q2$ are identical MOSFETs and $V_{GS1} = V_{GS2}$, $I_{D1} = 90\ \mu$A.

Recalling that the drain current for specified values of $V_T$ and $k$ is scaled by $W/L$, making the aspect ratios of $Q1$ and $Q2$ different permits $I_{D1}$ to differ from $I_R$ as given in Eq. (10-29) (Prob. 10-34):

$$\frac{I_{D1}}{I_R} = \frac{(W/L)_1}{(W/L)_2} \tag{10-29}$$

where $(W/L)_1$ and $(W/L)_2$ are the aspect ratios of $Q1$ and $Q2$, respectively.

The resistances $R_1$ and $R_2$ can be obtained from $V_{GG}$ and $R_G$ as

$$R_1 = R_G \frac{V_{DD}}{V_{GG}} = 100 \frac{28}{3} = 933 \text{ k}\Omega$$

$$R_2 = R_1 \frac{V_{GG}}{V_{DD} - V_{GG}}$$

$$= 933 \frac{3}{28 - 3} = 112 \text{ k}\Omega$$

Note that $R_G$ is selected arbitrarily at its minimum value. The drain resistance is obtained from the KVL equation for the drain-source loop. The voltage $V_{DS}$ is minimum when $I_D$ is maximum; thus

$$-V_{DD} + I_D R_D + V_{DS} + I_D R_S = 0$$

$$-28 + 6R_D + 8.5 + 6 \times 0.75 = 0 \qquad \text{and} \qquad R_D = 2.5 \text{ k}\Omega$$

Note that $R_D$ and $R_S$ cannot be specified independently. This situation is the same as for the BJT circuit in Fig. 10-12 described in the previous section. (See also Example 10-4.)

## 10-9 LINEAR ANALYSIS OF TRANSISTOR CIRCUITS

In the previous sections of this chapter we were concerned with biasing a transistor to establish a stable operating point. We now consider the response of transistor circuits to time-varying applied signals. In particular, we treat small-signal operation for which the transistors are assumed to behave linearly. Under these conditions, the signal component of the response is best obtained by using the small-signal (incremental) equivalent circuits of BJTs and FETs.

The low-frequency, small-signal equivalent circuit of a BJT was initially depicted in Fig. 3-33 and is repeated as Fig. 10-20 for convenience. Similarly, the values of the parameters in the model, given in Eqs. (3-28), (3-29), (3-33), and (3-34), are restated in Eqs. (10-32) through (10-35):

$$\beta_o = \left. \frac{\Delta i_C}{\Delta i_B} \right|_{V_{CEQ}} = \left. \frac{i_c}{i_b} \right|_{v_{ce}=0} \tag{10-32}$$

$$\beta_o = g_m r_\pi \qquad \text{or} \qquad r_\pi = \frac{\beta_o}{g_m} \tag{10-33}$$

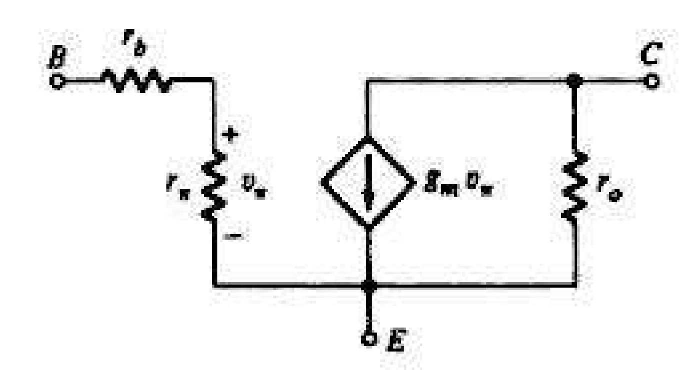



**FIGURE 10-20**
Low-frequency hybrid-$\pi$ equivalent circuit.

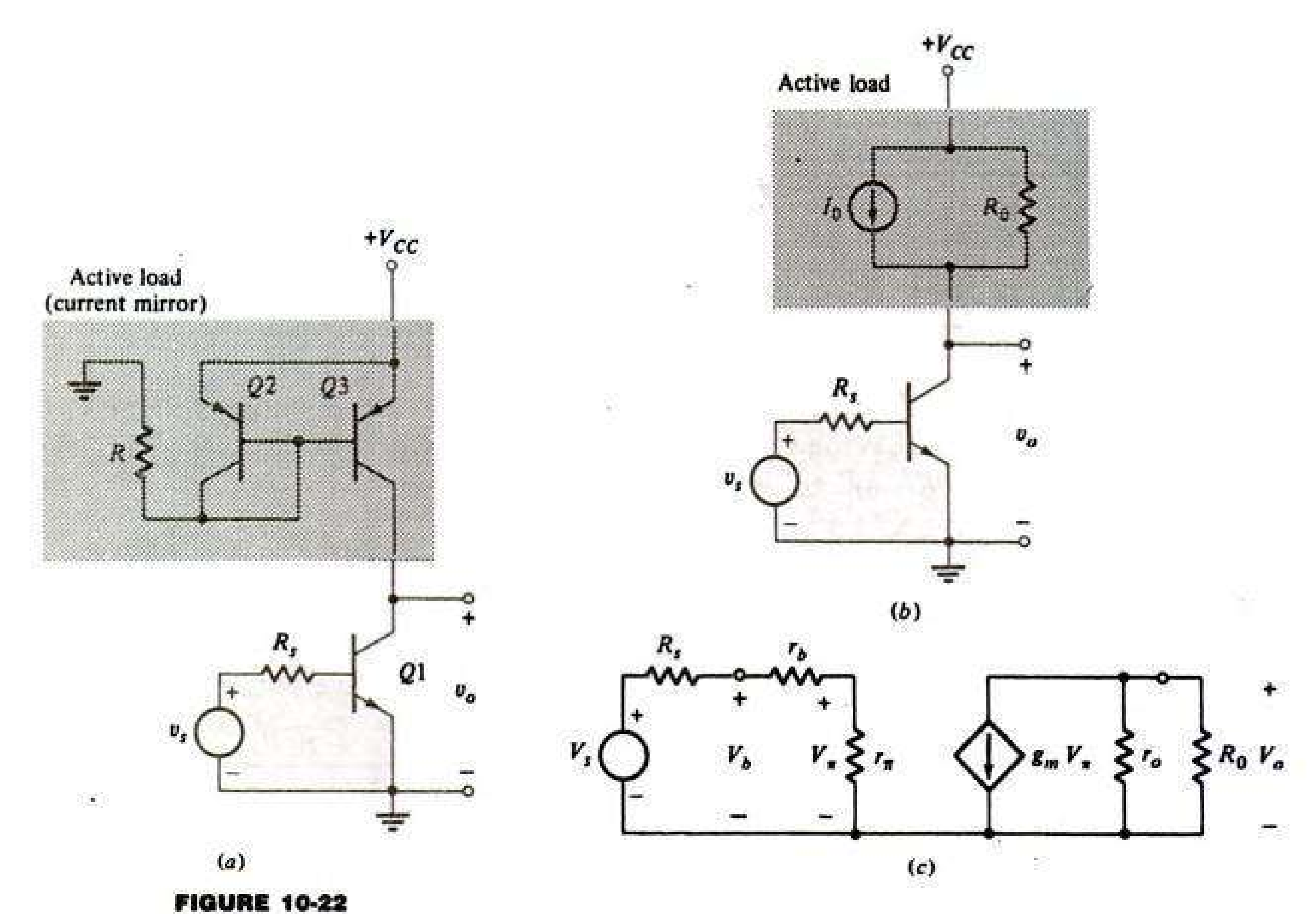



**FIGURE 10-22**
(*a*) A *pnp* current mirror used as a load in a common-emitter stage. (*b*) Equivalent representation of (*a*). (*c*) The small-signal equivalent circuit of the stage.

Sometimes, $V_o/V_b$ is identified as a voltage gain in the professional literature. This quantity is the voltage transfer ratio from input to output of the transistor and is often called the *transducer gain*. The transducer gain and $A_V$ and $A_I$ are related by

$$A_V = \frac{R_i}{R_i + R_s} \times \frac{V_o}{V_b} \quad \text{or} \quad \frac{V_o}{V_b} = -A_I \frac{R_L}{R_i} \tag{10-42a}$$

Because the transducer gain does not include the effect of the signal-source resistance $R_s$, it is generally less useful in the design of practical amplifiers.

**The Output Resistance** For the single-stage circuit, the output resistance $R_o$ is the resistance seen by the load $R_C$. By definition, $R_o$ is obtained by setting the source-voltage $V_s$ to zero and $R_C \rightarrow \infty$, applying a source $V_2$ to the output terminals, and measuring the current $I_2$ produced. Then $R_o \equiv V_2/I_2$. With $V_s = 0$, $I_b$ and $V_\pi$ are zero. Thus $I_2 = V_2/r_o$, and

$$R_o = \frac{V_2}{I_2} = r_o \tag{10-43}$$

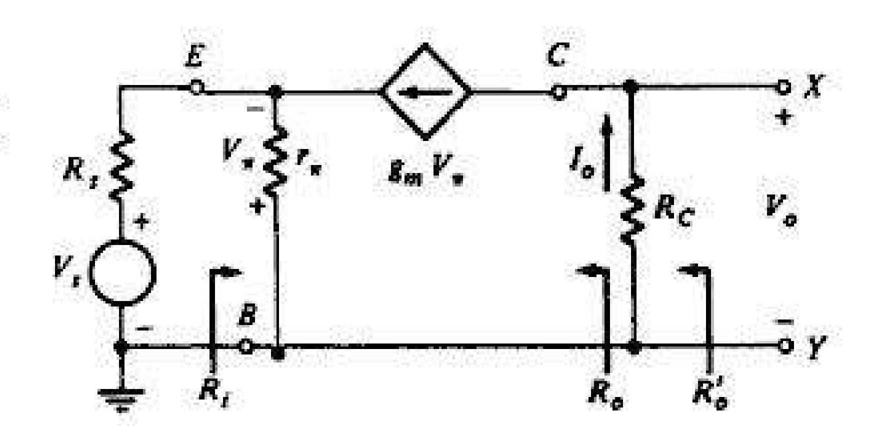



**FIGURE 10-24** Low-frequency equivalent circuit of the common-base stage.

## 10-12 THE COMMON-BASE AMPLIFIER

The circuit in Fig. 3-8 is a common-base amplifier stage if $V_{EE}$ and $R_E$ are replaced by a signal source $V_s$ having internal resistance $R_s$. The equivalent circuit is depicted in Fig. 10-24, and we observe that the use of the small-signal model and the results obtained are independent of whether an *npn* or a *pnp* transistor is under consideration. Applying the definitions of $A_I$, $R_i$, $A_V$, and $R_o$ in Sec. 10-10 to this circuit, we obtain the results given in the fourth column in Table 10-3. The verification of these formulas is left to the reader (Prob. 10-42).

## 10-13 COMPARISON OF BJT AMPLIFIER CONFIGURATIONS

Numerical values for $A_I$, $R_i$, $A_V$, $R_o$, and $R'_o$ for the three basic BJT amplifier configurations is given in Table 10-4 for $R_C = R_E = 1.5$ kΩ and $R_s = 0.60$ kΩ and a transistor having $\beta_o = 100$, $r_b = 50$ Ω, and $g_m = 0.10$ ℧. The value of $g_m$ corresponds to a BJT biased at $I_{CQ} = 2.5$ mA [Eq. (10-34)]. The value $r_o = 50$ kΩ is obtained for an Early voltage $V_A = 125$ V. For each configuration, three values of $A_I$, $R_i$, $A_V$, $R_o$, and $R'_o$ are given. The first column for each circuit is obtained by using the equations in Table 10-3*A* ($r_b = 0$, $r_o \to \infty$). Table 10-3*B* is applied to compute the remaining values, using $r_b = 50$ Ω, $r_o \to \infty$ in column 2 and $r_b = 0$, $r_o = 50$ kΩ in column 3 (for all three configurations). Comparison of the data for

**TABLE 10-4 Comparison of BJT Configurations***

| | CE | | | | CC | | | | CB | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Quantity | Approximate $r_b = 0$, $r_o \to \infty$ | | $r_b = 50$ Ω $r_o \to \infty$ | $r_b = 0$ $r_o = 50$ kΩ | Approximate $r_b = 0$, $r_o \to \infty$ | | $r_b = 50$ Ω $r_o \to \infty$ | $r_b = 0$ $r_o = 50$ kΩ | Approximate $r_b = 0$, $r_o \to \infty$ | | $r_b = 50$ Ω $r_o \to \infty$ | $r_b = 0$ $r_o = 50$ kΩ |
| $\|A_I\|$ | High | 100 | 100 | 97.1 | High | 101 | 101 | 98.1 | Low | 0.990 | 0.990 | 0.990 |
| $R_i$ | Medium | 1.00 kΩ | 1.05 kΩ | 1.00 kΩ | High | 153 kΩ | 153 kΩ | 147 kΩ | Low | 9.90 Ω | 10.4 Ω | 10.2 Ω |
| $\|A_V\|$ | High | 93.8 | 90.9 | 91.0 | Low | 0.990 | 0.989 | 0.989 | Low | 2.44 | 2.43 | 2.43 |
| $R_o$ | High | ∞ | ∞ | 50 kΩ | Low | 15.8 Ω | 16.3 Ω | 15.8 Ω | High | ∞ | ∞ | 1.93 MΩ |
| $R'_o$ | — | 1.50 kΩ | 1.50 kΩ | 1.46 kΩ | — | 15.6 Ω | 16.1 Ω | 15.6 Ω | — | 1.50 kΩ | 1.50 kΩ | 1.50 kΩ |

*Computed for $\beta_o = 100$, $g_m = 0.10$ ℧, $R_s = 0.60$ kΩ, and $R_C = R_E = 1.5$ kΩ.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

several hundred kilohms. The value of $g_m$ for a FET biased at $I_{DQ}$ is lower than the value of $g_m$ for a BJT biased at $I_{CQ} = I_{DQ}$. Hence, to achieve the same voltage gain as a BJT having a collector resistance $R_C$, the FET stage usually requires a drain resistance $R_D > R_C$. Because of the larger values of $R_D$ often used, the incremental output resistance $r_d$ cannot be neglected in the model. Because $r_d$ must be included in FET amplifier analysis, it is often convenient to use the model in Fig. 10-26*b*. The voltage source-resistance ($\mu V_{gs} - r_d$) combination in Fig. 10-26*b* is the voltage source equivalent of the $g_m V_{gs}$ current source in parallel with $r_d$ in Fig. 10-26*a*. The quantity $\mu$, called the *amplification factor*, is given by

$$\mu = g_m r_d \tag{10-64}$$

The open circuit between gate and source in the model[1] makes $I_g = 0$, so that it is meaningless to consider $A_I$ and $R_i$ for FET stages (usually). For most FET circuits, the voltage gain $A_V$ and the output resistance $R_o$ are the important quantities that describe amplifier performance.

The basic circuit we analyze is shown in Fig. 10-27*a*. Note that if the output voltage is measured from drain to ground (neutral, denoted by $N$), this stage is a *common-source stage with source resistance*. If $R_S = 0$, the stage is the standard *common-source (CS) stage* (Fig. 10-28*a*). Similarly, if the output is taken between source and ground, with $R_D = 0$, the circuit is a *common-drain (CD)* or *source-follower amplifier* in Fig. 10-28*b*. Inclusion of $R_D$ converts this stage to a common-drain stage with drain resistance.

**Analysis of the Generalized FET Amplifier Stage** The equivalent circuit of the generalized (basic) FET stage in Fig. 10-27*a* is given in Fig. 10-27*b*. Note that we use the equivalent circuit form in Fig. 10-26*b*. Application of KVL to the drain loop, assuming sinusoidal excitation, gives

$$I_d R_D + I_d r_d - \mu V_{gs} + I_d R_S = 0 \tag{10-65}$$

From Fig. 10-27*b* the voltage from $G$ to $S$ is given by

$$V_{gs} = V_i - I_d R_S \tag{10-66}$$

Combination of Eqs. (10-65) and (10-66) yields

$$I_d = \frac{\mu}{r_d + R_D + (1 + \mu)R_S} V_i \tag{10-67}$$

[1] A completely accurate model of the FET would include resistance $r_{dg}$ and $r_{sg}$ between drain and gate and source and gate, respectively. These resistors indicate current paths from source and drain to gate through the oxide layer in the MOSFET or through the reverse-biased *pn* junction in the JFET. Measurements (and theoretical analysis) indicate that $r_{dg}$ and $r_{sg}$ are greater than 10,000 MΩ ($\geq 10^{10}$ Ω), so it is reasonable to assume that these resistances are open circuits as shown.

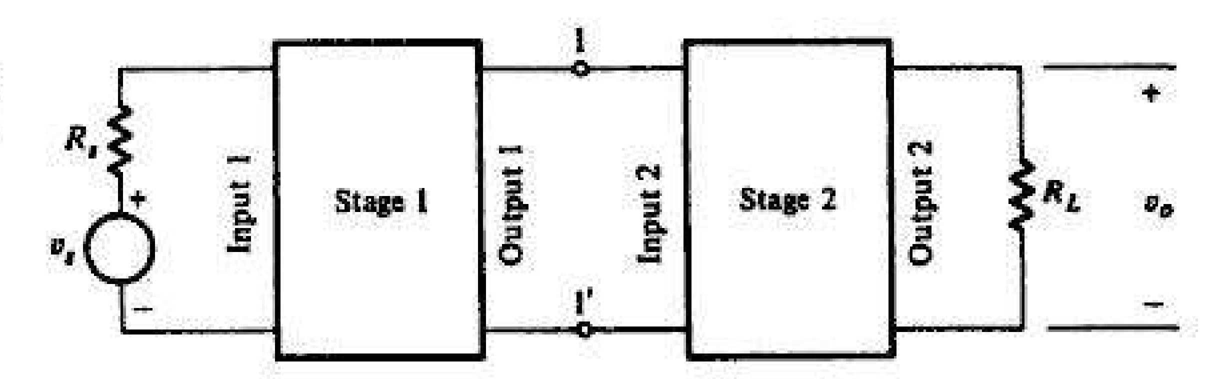



**FIGURE 10-29**
Pictorial representation of two cascaded stages.

not be the correct magnitudes for the intended application. To overcome these limitations, two or more stages can be connected in cascade; thus, the output of one stage is connected to the input of the next stage as seen in Fig. 10-29.

The analysis of the cascade configuration is based on the results obtained for single stages and is described as follows. We first obtain a Thévenin equivalent of the first stage at terminals 1–1′, that is, the portion of the circuit in Fig. 10-29 shown in Fig. 10-30*a*. For this single-stage amplifier, the output voltage (Thévenin voltage source) is $A_{V1}V_s$ and $R'_{o1}$ is the output resistance. This combination acts as the signal source and source resistance of the second stage as illustrated in Fig. 10-30*b*. For the stage illustrated in Fig. 10-30*b*, the output voltage is $A_{V2}$ times the input voltage or $V_o = A_{V1}A_{V2}V_s$ and

$$\frac{V_o}{V_s} = A_V = A_{V1}A_{V2} \tag{10-81}$$

The method leading to Eq. (10-81) is applicable to several stages by repeating the process. Note that if $A_{V1}$ and $A_{V2}$ are much greater than unity, $A_V$, the overall gain of the cascade, is much larger than the gain that can be achieved by either stage. We illustrate this in the following example.

*Example 10-7*

The cascaded amplifier in Fig. 10-31 consists of two common-emitter stages, one of which contains an emitter resistance and a common-collector stage. Note that biasing components are not shown and are assumed to have negligible effect. Transistor $Q1$ has $\beta_o = 100$ and $r_\pi = 1.0$ kΩ; transistors $Q2$ and $Q3$ have $\beta_o = 100$ and $r_\pi = 0.5$ kΩ. Determine the overall gain.

*Solution*

We first obtain the Thévenin equivalent of the first stage at terminals 1–1′. This is a common-emitter stage with an emitter resistance for which the voltage gain

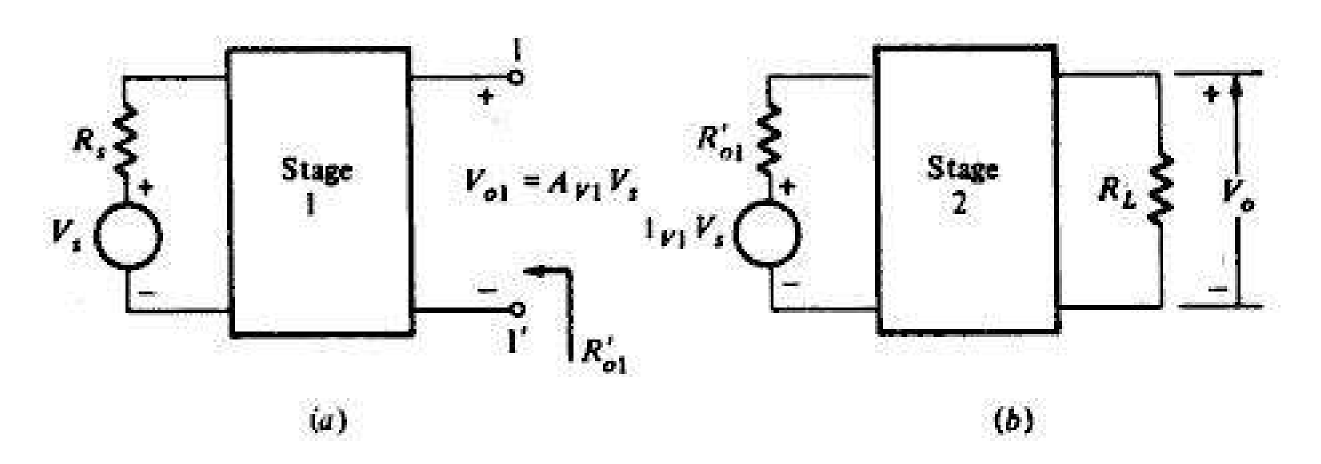
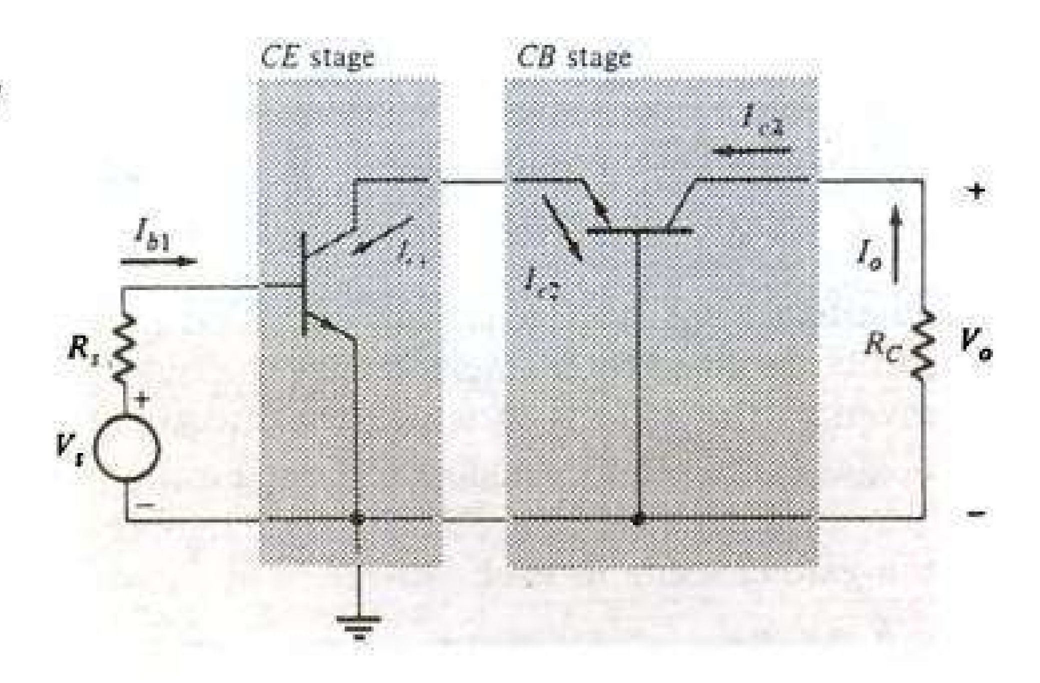



**FIGURE 10-30**
(*a*) The unloaded first stage of the amplifier in Fig. 10-29. (*b*) The second stage driven by the Thévenin equivalent of the first stage.

**FIGURE 10-35**
The cascode or *CE-CB* cascade.




that the input signal current in $Q2$ is the emitter current of $Q1$. Thus

$$I_{c2} = \beta_o I_{b2} = \beta_o(\beta_o + 1)I_{b1} \quad \text{and} \quad I_c = I_{c1} + I_{c2} = \beta_o I_{b1} + \beta_o(\beta_o + 1)I_{b1}$$

from which the current gain of the composite transistor $\beta_{oc}$ is

$$\beta_{oc} = \frac{I_c}{I_{b1}} = \beta_o(\beta_o + 2) \approx \beta_o^2 \tag{10-83}$$

for $\beta_o \gg 2$. For $\beta_o = 100$, $\beta_o^2 \approx 10^4$; clearly, the current gain is enhanced.

The Darlington transistor is most often used as an emitter follower because, as indicated in Table 10-3, the extremely high value of $\beta_{oc}$ makes $A_V$ virtually unity, $R_i$ extremely large, and $R_o$ extremely small.

**The *CC-CE* Connection** The circuit displayed in Fig. 10-34*b* is a *CC-CE* circuit and has properties similar to those of the Darlington pair. The current gain of the composite transistor is approximately $\beta_o^2$, since the emitter current in $Q1$ is the base current in $Q2$.

The *CC-CE* composite is preferred to the *CC-CC* configuration in constructing equivalent common-emitter stages. The advantage of the *CC-CE* configuration results from the fact that the collectors of $Q1$ and $Q2$ are not connected as in the Darlington pair. Through connection of the collectors (essentially in parallel), the output resistance ($r_o < \infty$) is reduced. Also, the frequency response of the *CC-CE* connection is superior to that for the *CC-CC* circuit.

**The *CE-CB* (Cascode) Connection** The primary use of the *cascode circuit* in Fig. 10-35 is to provide high voltage gain over a wider range of frequencies that can be achieved by a common-emitter stage. The frequency response of the *CE-CB* stage is discussed in Sec. 11-11. In Fig. 10-35, the current $I_{b1}$ is $V_s/(R_s + R_i)$ for the common-emitter stage or

$$I_{b1} = \frac{V_s}{R_s + r_{\pi 1}} \tag{10-84}$$

We note that $\beta_{o1} I_{b1} = I_{c1} = -I_{e2}$; for $\beta_{o2} \gg 1$, the current gain of the common-

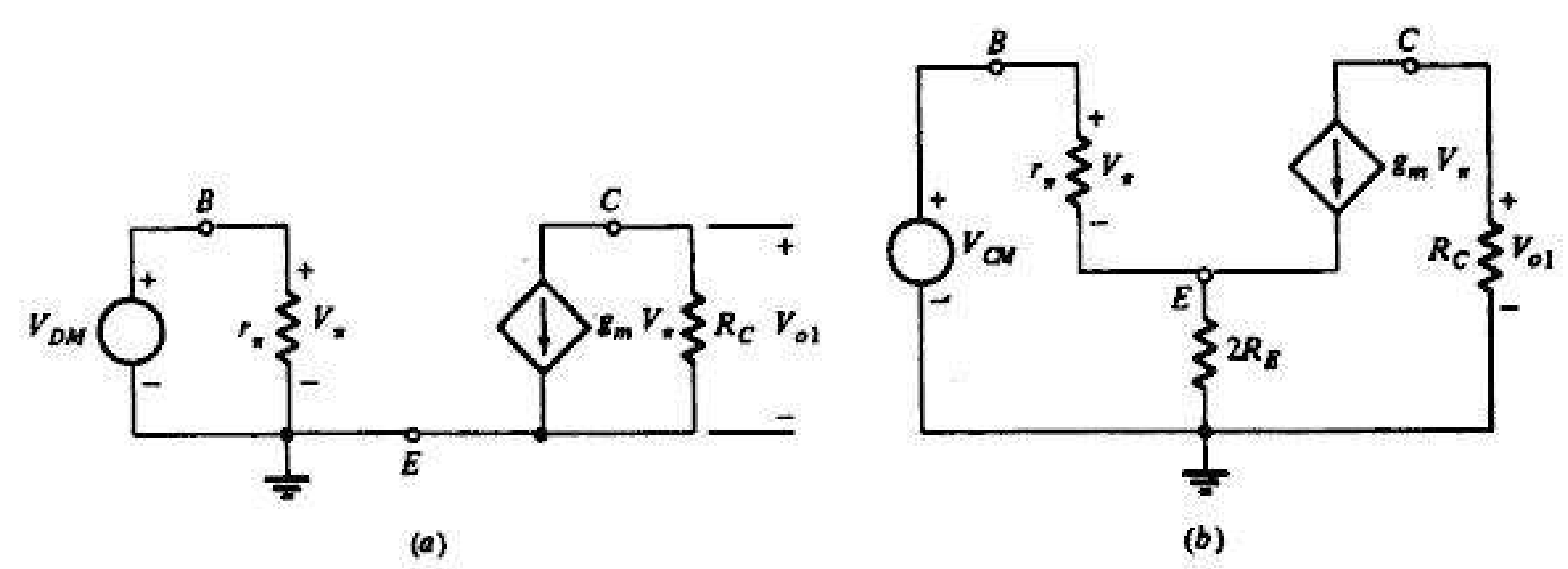



**FIGURE 10-38**
Small-signal models for (*a*) the differential mode and (*b*) the common mode.

$$A_{DM} = \frac{V_{o1}}{V_{DM}} = \frac{-\beta_o R_C}{r_\pi} = -g_m R_C \qquad (10\text{-}86)$$

For $V_{DM}$ positive, $V_{o1} = A_{DM}V_{DM}$ and, as seen in Eq. (10-86), $A_{DM}$ is negative, so that $V_{o1}$ is 180° out of phase with $V_{DM}$ ($V_{o1}$ is inverted). Because $Q_2$ is driven by $-V_{DM}$, $V_{o2} = -A_{DM}V_{DM}$ and $V_{o2}$ is in phase with $V_{DM}$ ($V_{o2}$ is noninverted).

**Common-Mode Gain $A_{CM}$** When a signal $V_{CM}$ is applied to both bases in Fig. 10-36 (common mode), the circuit in Fig. 10-37*b* is valid and results in the equivalent circuit of Fig. 10-38*b*. For this circuit, the gain $A_{CM}$ (from the $A_V$ entry in column 2 of Table 10-3*A*) is

$$A_{CM} = \frac{V_{o1}}{V_{CM}} = \frac{-\beta_o R_C}{2(\beta_o + 1)R_E + r_\pi} \qquad (10\text{-}87)$$

Equation (10-87), with $\beta_o \gg 1$ and division by $r_\pi$, reduces to

$$A_{CM} = \frac{-g_m R_C}{1 + 2g_m R_E} \approx -\frac{R_C}{2R_E} \qquad (10\text{-}88)$$

for $2g_m R_E \gg 1$. Because the same signal is applied to $Q1$ and $Q2$, both $V_{o1}$ and $V_{o2}$ are 180° out of phase with $V_{CM}$.

**The Common-Mode Rejection Ratio** The differential amplifier is primarily designed to amplify differential signals; hence we require $A_{DM} \gg A_{CM}$. A convenient measure of differential amplifier performance is the *common-mode rejection ratio CMRR*, defined as

$$CMRR \equiv \frac{A_{DM}}{A_{CM}} \qquad (10\text{-}89)$$

Combination of Eqs. (10-86) and (10-88) yields

$$CMRR = 1 + 2g_m R_E \approx 2g_m R_E \qquad (10\text{-}90)$$

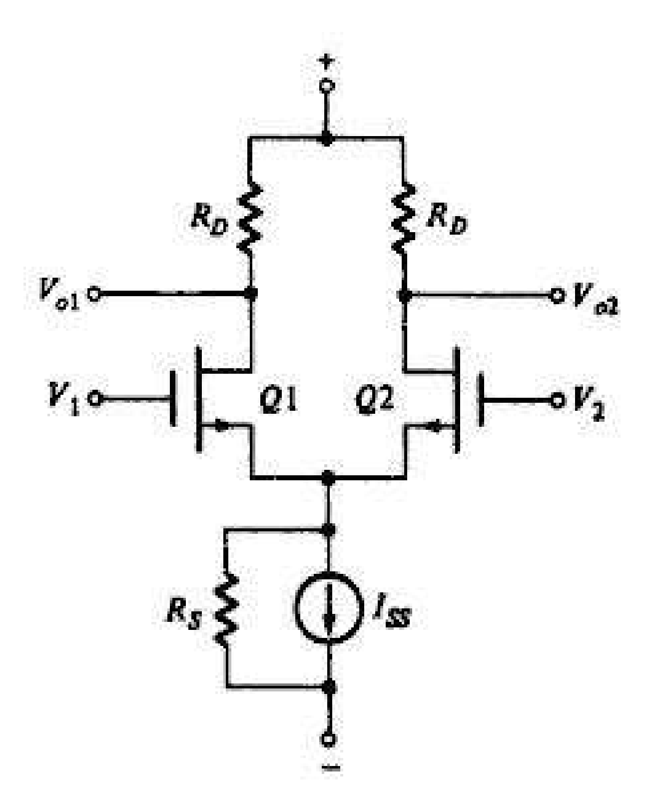



**FIGURE 10-40**
A source-coupled pair.

analysis that parallels that for the emitter-coupled pair described in Sec. 10-18. The differential-mode gain $A_{DM}$ is the gain of a common-source stage (Table 10-5), and for the common-source stage with source resistance, column 2 in Table 10-5 gives the common-mode gain $A_{CM}$. The common-mode rejection ratio *CMRR*, defined in Eq. (10-89), is

$$CMRR = 1 + \frac{2R_S(1 + \mu)}{r_d + R_D} \tag{10-96}$$

For $r_d \gg R_D$ and $\mu \gg 1$, Eq. (10-96) reduces to

$$CMRR = 1 + 2g_mR_S \approx 2g_mR_S \tag{10-97}$$

Note that the resistance $2R_S$ is derived from the half-circuit concept used to describe the differential amplifier.

Active loads, often depletion MOSFETs, are used to realize $R_D$ in Fig. 10-40. In fact, most MOSFET or JFET amplifiers fabricated on a chip employ active loads to both realize large resistance values and conserve chip area.

## 10-21 THE OPERATIONAL AMPLIFIER

The *operational amplifier*, or Op-Amp, is a direct-coupled, high-gain amplifier used to perform a wide variety of functions. It is often referred to as the *basic linear* (or more accurately *analog*) *integrated circuit* (IC), and many manufacturers package one to four identical units on a single chip. Although many Op-Amps comprise the cascade of a differential pair, common-emitter (source) stage, and emitter (source) follower, it is used widely as a single-stage amplifier. Our purpose in this section is to provide an introduction to the basic Op-Amp configurations, which we will encounter again in Chaps. 11 to 16. Chapter 14 is devoted exclusively to a more detailed examination of the internal design and several applications of Op-Amps. Many other circuits in which the Op-Amp is an integral component are treated in Part 4.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

and the current $I_B$ in $R_o$ is

$$I_B = \frac{-A_v V_i}{R_o} \tag{10-107}$$

Combining Eqs. (10-106), (10-107), and (10-103) with $V_o = 0$, we obtain

$$I_{sc} = I_A + I_B = \frac{-A_v(R_2 + R_o)V_1}{R_1(1 + A_v) + R_2 + R_o} \tag{10-108}$$

from which

$$R_T = \frac{(R_1 + R_2)R_o}{R_1(1 + A_v) + R_2 + R_o} = \frac{[(R_1 + R_2)R_o]/(R_1 + R_2 + R_o)}{1 + [A_v R_1/(R_1 + R_2 + R_o)]} \tag{10-109}$$

The value of $R_T$ is significantly lower than $R_o$. In the right-hand form of Eq. (10-109), the numerator is $R_o \| (R_1 + R_2)$, which is smaller than $R_o$. This resistance is divided by a large (for $A_v \gg 1$) positive term, and, hence, $R_T \to 0$ as $A_v \to \infty$.

***Example 10-9*** An inverting Op-Amp stage is designed with $R_1 = 5$ kΩ, $R_2 = 10$ kΩ, and $R_L = 100$ Ω. The Op-Amp has $A_v = 5 \times 10^4$, $R_o \doteq 500$ Ω, and $R_i \to \infty$. (*a*) Determine the voltage across $R_L$ for an rms input signal of 1.5 V. (*b*) Repeat part *a*, considering that the Op-Amp is ideal.

***Solution*** (*a*) The Thévenin equivalent of the stage is given by Eqs. (10-105) and (10-109). Evaluation gives

$$A_v V_s = \frac{-5 \times 10^4 \times 10 + 0.5}{5(1 + 5 \times 10^4) + 10 + 0.5} \times 1.5 = -2.9998 \text{ V}$$

$$R_T = \frac{(5 + 10)0.5}{5(1 + 5 \times 10^4) + 10 + 0.5} = 0.027 \ \Omega$$

The voltage across $R_L$, using the Thévenin equivalent, is

$$V_o = \frac{R_L}{R_L + R_o} A_V V_1 = \frac{100}{100 + 0.027} \times (-2.9998) = -2.999 \approx -3.0 \text{ V}$$

(*b*) For the ideal Op-Amp, the voltage across $R_L$ is simply $A_V V_1$ or $V_o = -(R_2/R_1)\, V_1 = -\frac{10}{5} \times 1.5 = -3.0$ V.

The results indicate that for typical values of $A_V$ and $R_o$, the difference between actual and ideal values is negligible.

**The Practical Noninverting Stage** The practical noninverting Op-Amp stage (shown in Fig. 10-45) is analyzed in an identical manner as is the inverting stage. The value of $A_V$ is given in Eq. (10-110) and that for $R_T$ in Eq. (10-109), assuming $R_i \to \infty$:

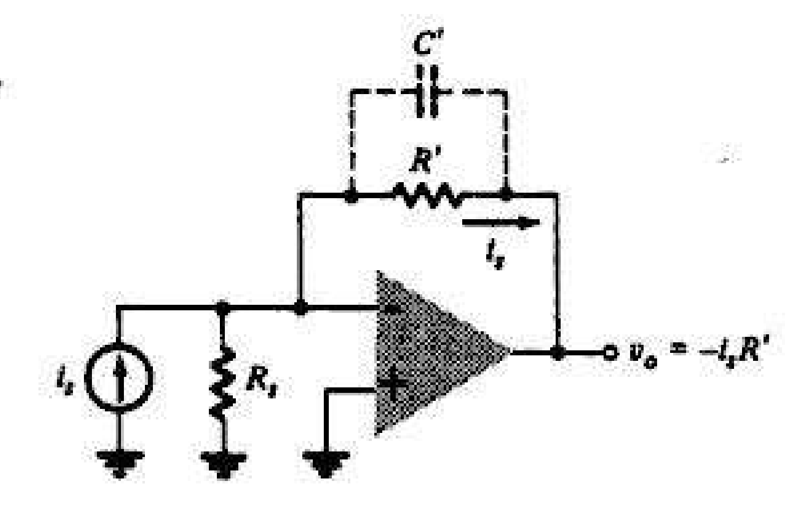



**FIGURE 10-49** Current-to-voltage converter.

## Integrators

If, in Fig. 10-42*a*, the resistance $R_2$ is replaced by a capacitor $C$, as displayed in Fig. 10-50, the circuit behaves as an *integrator*. In Fig. 10-50, $i_1 = v_s/R_1$ and $i_C = C(dv_o/dt)$; since no current enters the Op-Amp, $i_1 = -i_C$. Hence

$$\frac{v_s}{R_1} = -C\frac{dv_o}{dt}$$

or, on integrating and solving for $v_o$

$$v_o = -\frac{1}{R_1 C}\int v_s\,dt \tag{10-117}$$

The amplifier therefore provides an output voltage proportional to the integral of the input voltage.

If the input voltage is a constant, $v_s = V$, then the output will be a ramp, $v_o = -Vt/R_1C$. Such an integrator makes an excellent sweep circuit for a cathode-ray-tube oscilloscope, and is called a *Miller integrator*, or *Miller sweep*. The circuit in Fig. 10-50 is an ideal integrator. Practical integrators are treated in detail in Sec. 16-6.

## REFERENCES

1 Ghausi, M. S.: "Electron Devices and Circuits: Discrete and Integrated," Holt, Rinehart and Winston, Inc., New York, 1985.

2 Sedra, A. S., and K. C. Smith: "Microelectronic Circuits," Holt, Rinehart and Winston, Inc., New York, 1981.

3 Schilling, D., and C. Belove: "Electronic Circuits—Discrete and Integrated," McGraw-Hill Book Company, New York, 1979.

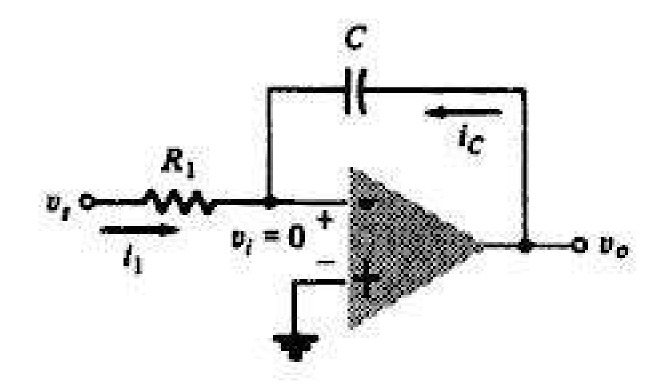



**FIGURE 10-50** Op-Amp integrator circuit.

# Chapter 11
# FREQUENCY RESPONSE OF AMPLIFIERS

The signals utilized in many electronic systems require amplification with a minimum of distortion. Under these circumstances, the active devices involved must operate linearly; thus, small-signal conditions must apply. The first step in the analysis of such circuits is the use of a linear model to replace the actual circuit. Thereafter it is a matter of circuit analysis to determine the transmission characteristics of the linear network.

In Chap. 10, we focused on the low-frequency behavior of amplifier stages. To do so we considered that both the internal capacitances of the transistors and, where employed, the external coupling and bypass capacitors had negligible effect on performance. However, amplifiers are required to operate over a broad frequency range. The low end of this range may be dc (direct-coupled stages) or a few hertz, and the high end may extend to several tens of megahertz. The study of such wideband amplifiers was given impetus by the need to amplify the pulses that occur in communications systems such as television and radar. In the analysis of amplifiers over so wide a frequency band, the capacitances, heretofore neglected, must be included. Because capacitive reactance varies with frequency, the transmission characteristics of the linear models are frequency-dependent. Thus amplifier gain depends on the frequency of the input signal which can result in output signals that are frequency distorted.

In this chapter we consider how a low-level input signal that contains many frequency components from zero (dc) to a few megahertz can be amplified with a minimum of distortion.

In addressing this question, we first investigate the response of single-stage BJT and FET amplifiers. Multistage amplifiers are treated by relating the overall response to the frequency response of the component stages. Approximate methods for evaluating the frequency response, useful in amplifier design, are developed.

Integrated-circuit amplifiers, which are invariably direct-coupled, are limited only at high frequencies because of internal transistor capacitances ($C_\pi$ and $C_\mu$

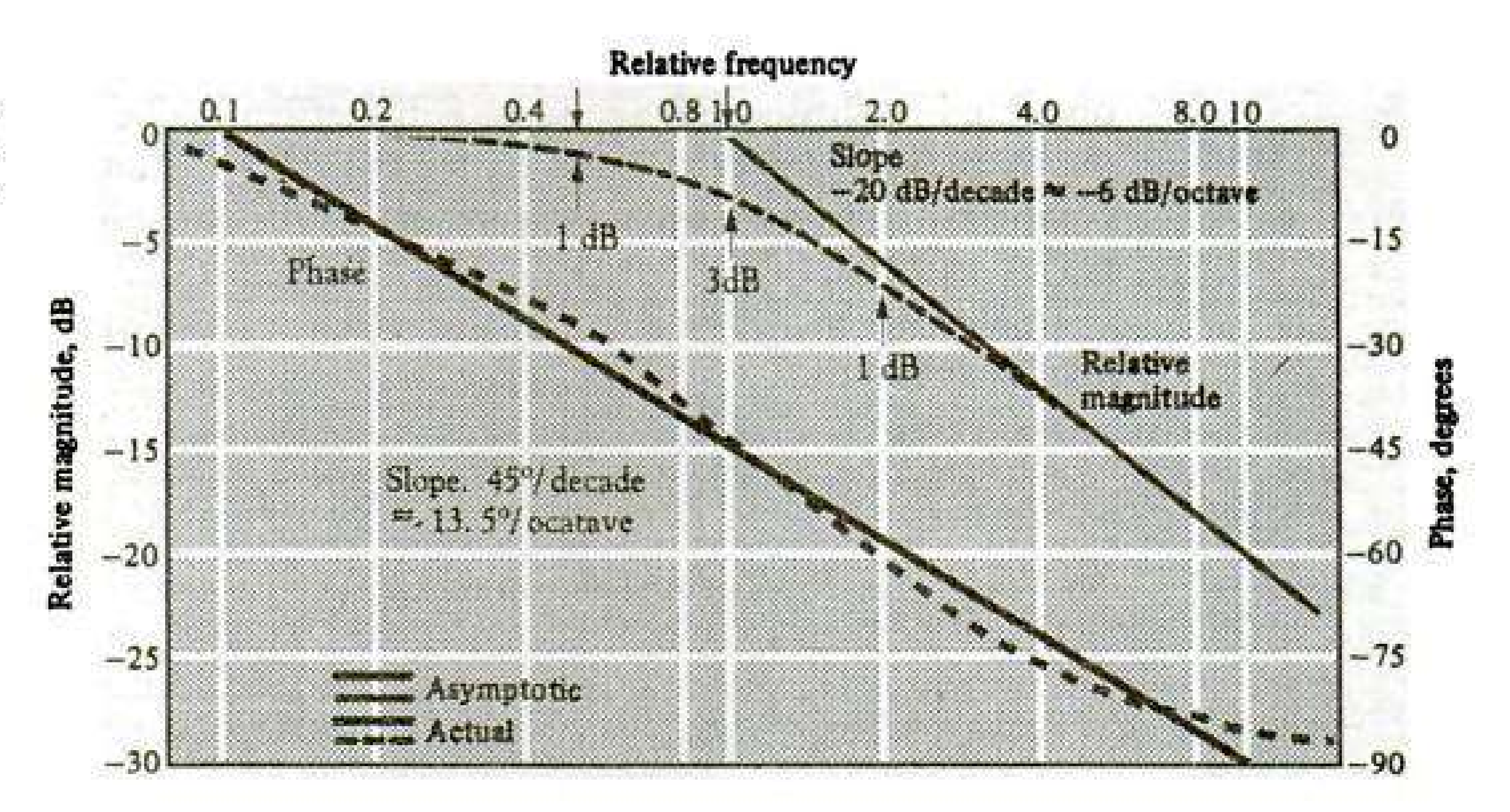



**FIGURE 11-3** Normalized Bode diagram of the transfer function of the circuit in Fig. 11-2*b*.

and $f_H$ is often referred to as the *high*, or *upper 3-dB frequency*. As power is proportional to the square of the voltage, $A_{VH} = A_{VO}/\sqrt{2}$ corresponds to a power level at $f = f_H$ equal to one-half the power at $f = 0$. Thus $f_H$ is also referred to as the *upper (high) half-power frequency*.

The frequency-response characteristic given in Eq. (11-5) is displayed in the Bode diagram in Fig. 11-3. The dashed curves indicate actual values, and the solid curves are for the asymptotic Bode diagram. Note that the frequency axis in Fig. 11-3 is normalized to $f/f_H$ and that the gain is normalized to $A_{VH}/A_{VO}$. This circuit behaves as a *low-pass system*, for, as seen in the asymptotic Bode diagram, frequencies below $f_H$ are transmitted with minimum attenuation and frequencies above $f_H$ are attenuated.

For typical parameter values encountered ($C_M$ = 40 pF, $R_i = R_s$ = 1 kΩ) in practical amplifier stages, $f_H \simeq 8$ MHz. Obviously, increasing (decreasing) $R_i$ and $R_s$ (or $C_M$) results in decreasing (increasing) $f_H$.

**Low-Frequency Response** The circuit in Fig. 11-4 represents an amplifier stage in which $1/sC_C$ represents the impedance of the external coupling capacitor ($C_B$ in Fig. 10-12 or $C_G$ for an FET stage). For Fig. 11-4

$$V_1 = \frac{R_i V_s}{R_s + R_i + 1/sC_C}$$

and, from $V_o = -g_m R_L V_1$

$$A_{VL}(s) = \frac{V_o}{V_1} = \frac{-g_m R_L R_i}{R_i + R_s + 1/sC_C} \qquad (11\text{-}5)$$

Equation (11-5) can be recast as

$$A_{VL}(s) = \frac{-g_m R_L R_i/(R_s + R_i)}{1 + 1/\,sC_C(R_s + R_i)} = \frac{|A_{VO}|}{1 + \omega_L/s} = |A_{VO}|\,\frac{s/\omega_L}{1 + s/\omega_L} \qquad (11\text{-}6)$$

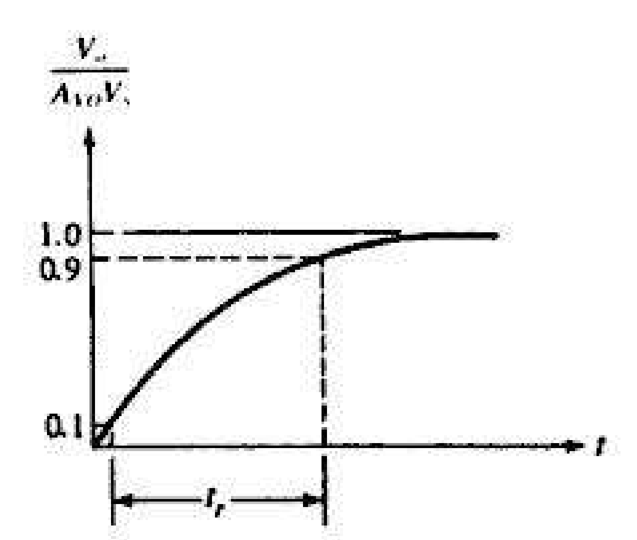



**FIGURE 11-8**
Normalized step response of the circuit in Fig. 11-2*a*.

the distortion of the flat portion of the step. We should, of course, expect such relationships, since the high-frequency response measures essentially the ability of the amplifier to respond faithfully to rapid variations in signal, whereas the low-frequency response measures the fidelity of the amplifier for slowly varying signals. An important feature of a step is that it is a combination of the most abrupt voltage change possible and the slowest possible voltage variation.

**Rise Time**

The response of the low-pass circuit in Fig. 11-2 to a step input of amplitude $V_s$ is exponential with a time constant $1/\omega_H$. Since the capacitor voltage cannot change instantaneously, the output starts from zero and approaches its steady-state value $A_{VO}V_s$. The output is given by

$$v_o = A_{VO}V_s\,(1 - \epsilon^{-\omega_H t}) \tag{11-8}$$

The time required for $v_o$ to reach one-tenth of its final value is readily found to be $0.1/\omega_H$, and the time needed to reach nine-tenths its final value is $2.3/\omega_H$. Note that the time constant is $C_M$ multiplied by the equivalent resistance $R_s \parallel R_i$. The difference between these two values is called the *rise time* $t_r$ of the circuit and is shown in Fig. 11-8. The time $t_r$ is an indication of how fast the amplifier can respond to a discontinuity in the input voltage. We have

$$t_r = \frac{2.2}{\omega_H} = \frac{2.2}{2\pi f_H} = \frac{0.35}{f_H} \tag{11-9}$$

Note that the rise time is inversely proportional to the upper 3-dB frequency. For an amplifier with 1-MHz bandpass, $t_r = 0.35\ \mu s$.

The relationship between $t_r$ and $f_H$ in Eq. (11-9) is exact (to three significant figures) for a single-pole circuit. However, Eq. (11-9) is a good approximation (within 3 or 4 percent) for multiple-pole low-pass circuits.

Consider a pulse of width $T_p$. What must be the upper 3-dB frequency $f_H$ of an amplifier if the signal is to be amplified without excessive distortion? A reasonable answer to this question is: *Choose $f_H$ equal to or greater than the reciprocal of the pulse width $T_p$*. For $f_H = 1/T_p$, the output waveform (color) in Fig. 11-9 is the response to the input pulse indicated.

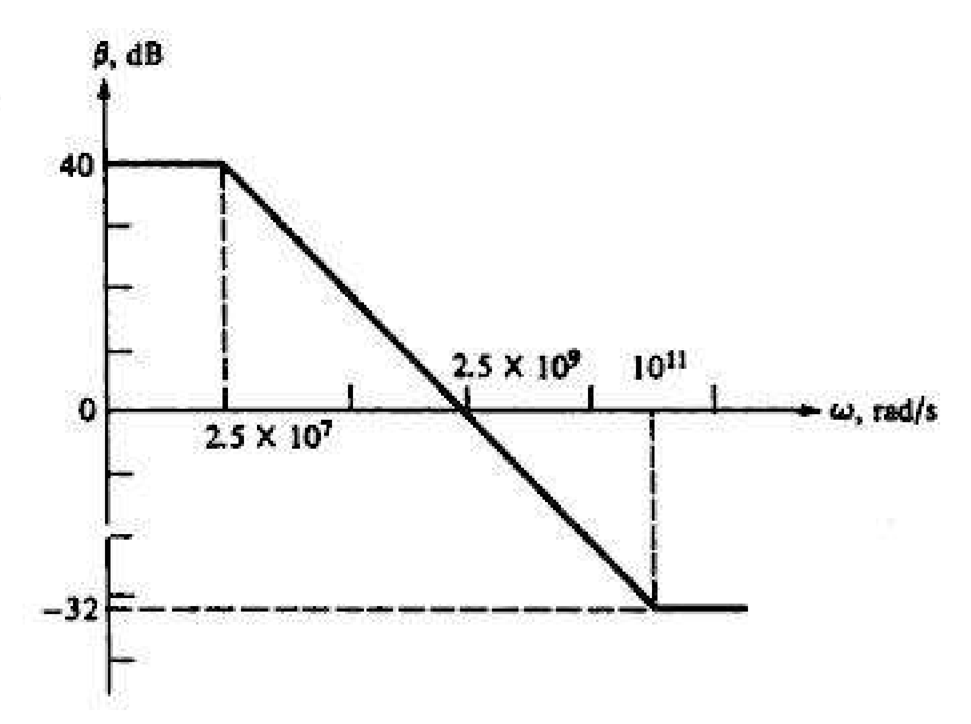



**FIGURE 11-12** Asymptotic Bode diagram for $|\beta(j\omega)|$.

**The Parameter $f_T$** In Example 11-1 the use of the asymptotic characteristic gives a highly accurate result for the unity gain frequency because of the wide separation of $\omega_\beta$ and $\omega_z$. To build an amplifier with current gain greater than unity, it is evident that operating frequencies must be less than $f_T$. Consequently, for $\omega \leq \omega_T$, Eq. (11-16) can be approximated by the single-pole function

$$\beta(s) = \frac{\beta_o}{1 + s/\omega_\beta} \tag{11-17}$$

The approximation in Eq. (11-17) is equivalent to stating that the current $I_\mu$ in $C_\mu$ is a negligible component of $I_o$ ($I_\mu \ll g_m V_\pi$). The asymptotic Bode diagram for this function is identical to that in Fig. 11-12 for the frequency range $\omega < \omega_z$.

To determine $f_T$, *the frequency at which the common-emitter short-circuit current gain has unit magnitude*, Eq. (11-17) is used. Hence

$$|\beta(j\omega_T)| = 1 = \frac{\beta_o}{\sqrt{1 + (\omega_T/\omega_\beta)^2}}$$

which, for $\beta_o^2 \gg 1$, gives

$$\omega_T = \beta_o \omega_\beta \qquad \text{or} \qquad f_T = \beta_o f_\beta \tag{11-18}$$

Substitution of $\omega_\beta$ in Eq. (11-16) into Eq. (11-18) yields

$$f_T = \frac{\beta_o}{2\pi\, r_\pi (C_\pi + C_\mu)} = \frac{g_m}{2\pi(C_\pi + C_\mu)} \tag{11-19}$$

Note that for $C_\pi \gg C_\mu$, $f_T \approx g_m/2\pi C_\pi$. The parameter $f_T$, as with other BJT parameters, depends on the operating conditions of the device. Typically, $f_T$ varies with quiescent collector current as displayed in Fig. 11-13.

Since $f_T \approx \beta_o f_\beta$, this parameter may be given a second interpretation. It represents the *short-circuit current gain-bandwidth product*; that is, for the *CE* configuration with the output shorted, $f_T$ is the product of the low-frequency

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

The importance of Eq. (11-24) is that we can approximate the pole locations by knowing the coefficients $a_1$, $a_2$, and $a_3$ in $A_H(s)$. Furthermore, the dominant-pole approximation gives the value of the 3-dB frequency $f_H$ as

$$f_H \simeq \frac{p_1}{2\pi} = \frac{1}{2\pi a_1} \tag{11-25}$$

The form of Eq. (11-24), $p_k = a_{k-1}/a_k$, applies to a transfer function of $n$ real poles. In relation to amplifier response, however, we are interested in $p_1$ and $p_2$ ($a_1$ and $a_2$) only. First, $p_1$ determines the approximate value of $f_H$, and second, the separation between $p_1$ and $p_2$ indicates the degree to which the dominant-pole approximation is valid. For most of the transfer functions encountered, $p_2/p_1 = a_1^2/a_2 \geq 8$ gives $f_H$ within 10 percent and $p_1$ within 20 percent of their actual values. As $p_2/p_1$ increases, the error between actual and approximate values decreases. Note that the dominant-pole approximation *always* results in values of $f_H$ and $p_1$ that are less than the corresponding actual values.

Of greater importance than the numerical convenience afforded by Eqs. (11-24) and (11-25) is that the coefficients $a_1$, $a_2$ and so on can be determined by circuit time constants as described in Sec. 11-9. This permits the circuit designer to correlate overall response with the particular components (stages) which produce the response.

The student should note the following three limitations of the dominant-pole method:

1. It is valid only for transfer functions with real poles.
2. Any (all) zeros in the transfer function must be at least two octaves beyond the dominant pole.
3. The representation of the transfer function by a single dominant pole does not give accurate results for the phase characteristic.

## 11-5 THE HIGH-FREQUENCY RESPONSE OF A COMMON-EMITTER STAGE

The equivalent circuit used to evaluate the high-frequency performance of the basic common-emitter stage shown in Fig. 10-21 is displayed in Fig. 11-14. Typical numerical values of transistor parameters and circuit components are also indicated in Fig. 11-14. Since the circuit has two independent capacitors, the

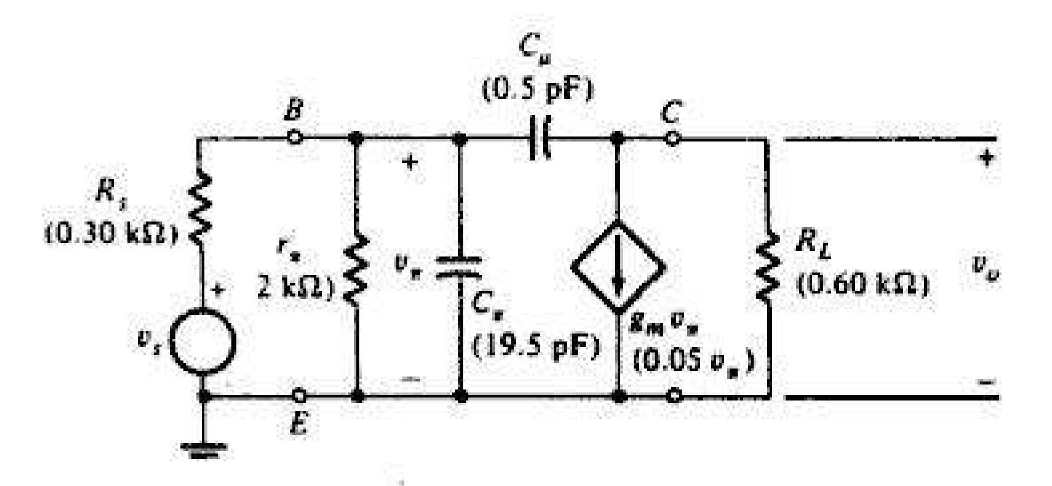



**FIGURE 11-14**
High-frequency equivalent circuit of the basic common-emitter stage (Fig. 10-21*a*). The hybrid-$\pi$ model for the BJT is shown in blue.

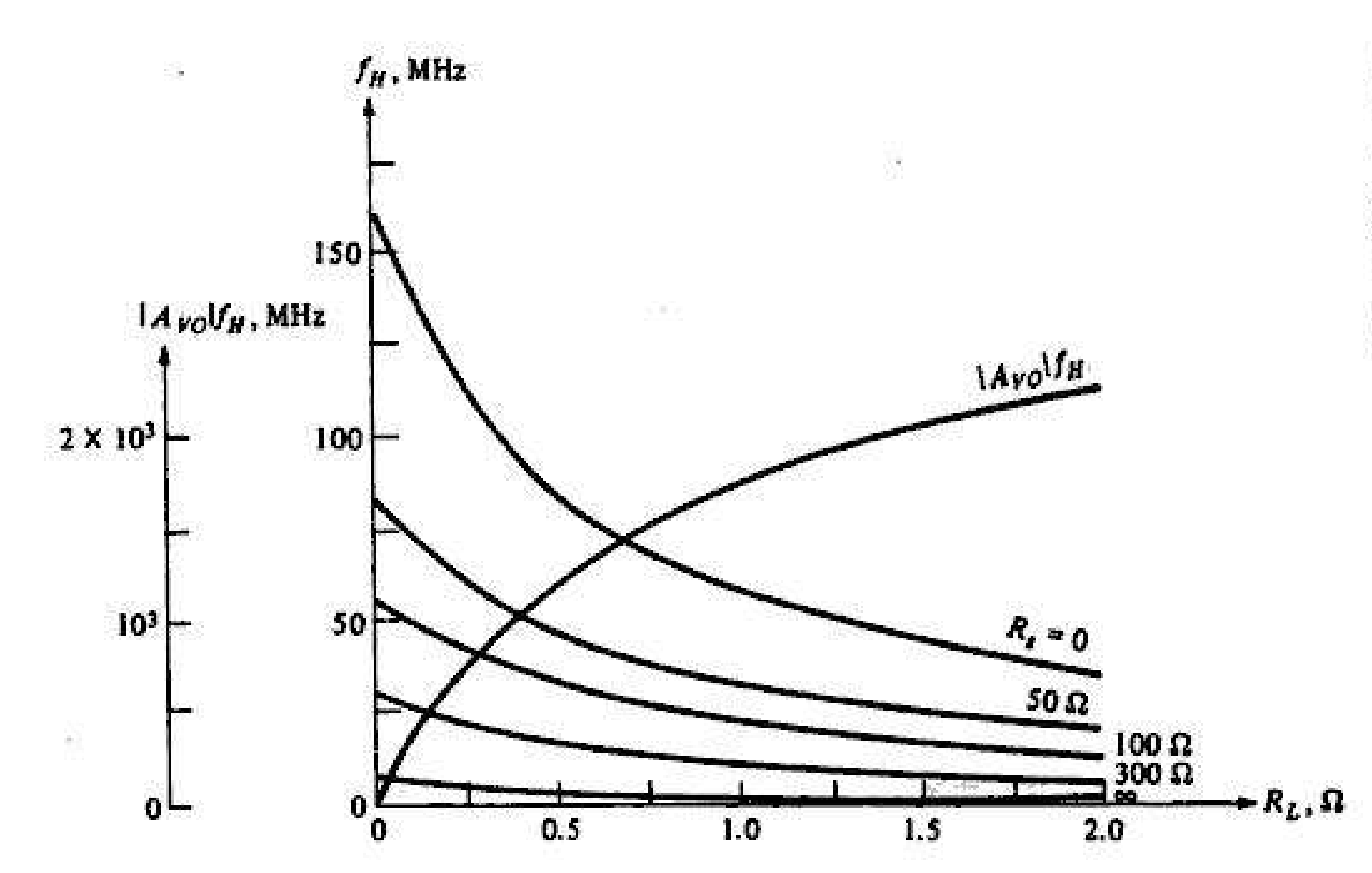



**FIGURE 11-16**
Bandwidth $f_H$ as a function of $R_L$, with source resistance $R_s$ as a parameter, for a single-stage common-emitter amplifier. Equation (11-31) is used to compute $f_H$ using the numerical values given in Fig. 11-14.

11-16 for $R_s = 0$ corresponds to ideal-voltage-source drive.[1] We note that the bandwidth is highest for any value of $R_L$ for the lowest $R_s$. The voltage gain-bandwidth product increases with increasing $R_L$ and decreases with increasing $R_s$. Even if we know the gain-bandwidth product at a particular $R_s$ and $R_L$, we cannot use the product to determine the improvement in bandwidth corresponding to a sacrifice in gain. If we change the gain by changing $R_L$ or $R_s$ or both, in general, the gain-bandwidth product will no longer be the same as it had been.

## 11-7 THE COMMON-SOURCE STAGE AT HIGH FREQUENCY

The high-frequency analysis of the common-source stage parallels that for the common-emitter stage discussed in the previous section. Note the similarity of the high-frequency model of the common-source stage in Fig. 11-17 with Fig. 11-14 for the common-emitter stage. If $r_b = 0$ and $r_\pi \rightarrow \infty$ in Fig. 11-14, the only difference between the two circuits is the existence of a third capacitance $C_{ds}$ at the output in Fig. 11-17. Although there are three capacitances, these are not independent because they form a loop (Sec. 11-4). Thus the transfer function $A_{VH}(s) = V_o/V_s$ has only two poles and, by analogy with Fig. 11-14, one finite zero.

The nodal equations for the circuit in Fig. 11-17 are

$$\frac{V_s}{R_s} = V_{gs}\left(\frac{1}{R_s} + sC_{gs} + sC_{gd}\right) - sC_{gd}V_o \qquad (11\text{-}34a)$$

---

[1]Note that for $R_s = 0$, the value of $f_H$ shown is in error because the time constant $C_\mu R_L$ is comparable to $R_\pi^o C_M$. However, the form of the variation in $f_H$ displayed is indicative of the physical situation that exists.

Similarly

$$p_2 = \frac{a_1}{a_2} = \frac{0.566}{0.0573} \times 10^9 = 9.87 \times 10^9 \text{ rad/s}$$

or $f_2 = p_2/2\pi = 1.57$ GHz.

The ratio $p_2/p_1 = 5.58$ and the dominant-pole approximation does not give particularly accurate results. Solving for the poles from Eq. (11-40) yields $p_1 = 2.30 \times 10^9$ rad/s and $p_2 = 7.57 \times 10^9$ rad/s. With these values, $f_H =$ 339 MHz.

We observe that $f_H$, the pole frequencies, and the zero frequency are all of the order of magnitude of $f_T$. Because these values are at the frequency limit for which the equivalent circuit is valid, the numerical values are subject to question. However, they all display the approximate magnitudes obtained from computer simulations.

Comparison of $f_H$ for the common-collector stage with $f_H = 16.9$ MHz for the common-emitter stage in Sec. 11-6 indicates that the common-collector stage has a considerably larger bandwidth than does the common-emitter stage. In reality, a common-emitter stage having $R_C = 1.5$ kΩ, driven from a source $R_s = 0.6$ kΩ and using the same transistor, has $f_H \approx 4.37$ MHz. Thus we conclude that when a common-collector stage is driven by (drives) a common-emitter stage, the value of $f_H$ for the cascade is simply that of the common-emitter stage.

**The Output Impedance $Z_o$** The high-frequency output impedances $Z_o$ and $Z_o'$ are obtained from the Thévenin equivalent of the stage. The open-circuit voltage, measured across $R_E$, is simply $V_o = A_{VH}V_s$. The current $I_L$ is $V_o/R_E$ and, for $R_E = 0$, gives the short-circuit current $I_{sc}$ as

$$I_{sc} = \frac{(\beta_o + 1)V_s}{R_s + r_\pi} \frac{1 + sr_\pi C_\pi/(1 + \beta_o)}{1 + sR_s r_\pi (C_\pi + C_\mu)/(R_s + r_\pi)} \tag{11-43}$$

The Thévenin impedance $Z_o' = V_o/I_{sc}$ and is

$$Z_o' = \frac{R_E\,[(R_s + r_\pi)/(1 + \beta_o)]}{R_E + [(R_s + r_\pi)/(1 + \beta_o)]} \frac{1 + sr_\pi R_s\,(C_\pi + C_\mu)/(R_s + r_\pi)}{1 + a_1 s + a_2 s^2} \tag{11-44}$$

where $a_1$ and $a_2$ are as given in Eq. (11-41).

We observe in Fig. 11-19 that $Z_o' = Z_o \parallel R_E$ and if $R_E \to \infty$, $Z_o' = Z_o$. Hence from Eq. (11-44), we obtain

$$Z_o = \frac{R_s + r_\pi}{1 + \beta_o} \frac{1 + sr_\pi R_s\,(C_\pi + C_\mu)/(R_s + r_\pi)}{[1 + sr_\pi C_\pi/(1 + \beta_o)](1 + sR_s C_\mu)} \tag{11-45}$$

We can identify $(R_s + r_\pi)/(1 + \beta_o)$ as the low-frequency output resistance $R_o$ (Table 10-3). Thus $Z_o$ is of the form

$$Z_o = R_o \frac{(1 + s/z_1)}{(1 + s/p_1)(1 + s/p_2)} \tag{11-46}$$

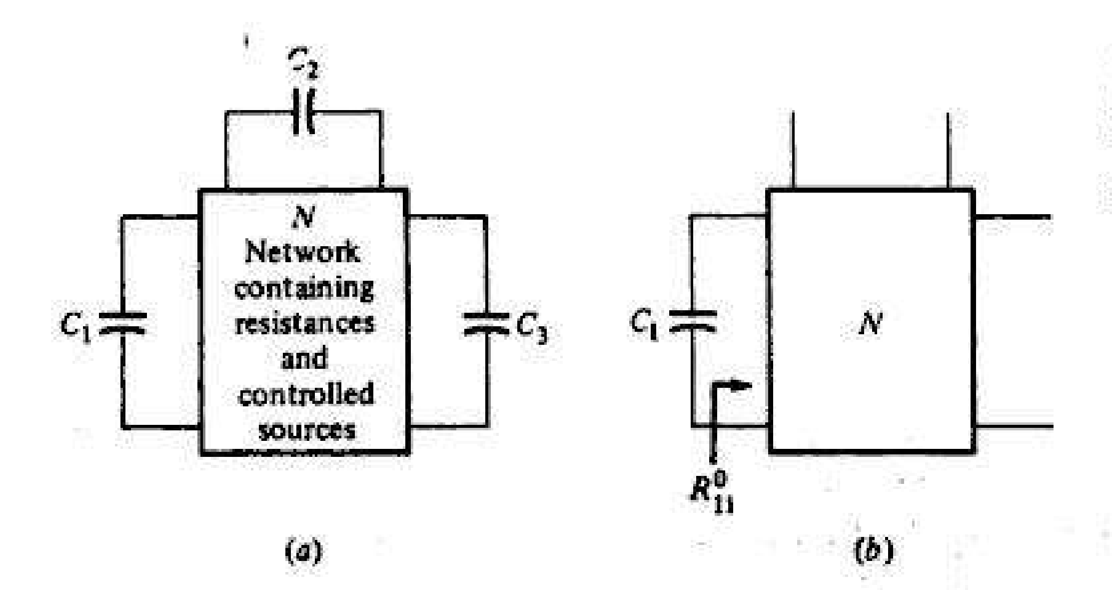



**FIGURE 11-23**
(*a*) Network containing three capacitors. (*b*) The network in (*a*) with $C_2$ and $C_3$ open-circuited. The zero-frequency resistance $R_{11}^0$ is defined by this configuration.

## 11-9 THE TIME-CONSTANT METHOD OF OBTAINING THE RESPONSE

The form of Eqs. (11-28), (11-35), (11-40), and (11-53) in which the coefficient $a_1$ is the sum of time constants and $a_2$ is the product of time constants is not unique to simple common-emitter, common-source, common-collector, and common-drain stages. These coefficients in the characteristic polynomial of any linear system containing resistors, capacitors, and controlled sources can *always* be expressed in this form. In this section we describe a method for obtaining these coefficients by direct determination of the equivalent resistances needed to evaluate the time constants.[1]

**The Coefficient $a_1$** Consider the network in Fig. 11-23*a* in which the portion of the circuit $N$ within the rectangle contains only resistors and controlled sources. With three independent capacitors indicated, the transfer function has three poles, so that its denominator is expressed as given in Eq. (11-22). The coefficient $a_1$ can be written as

$$a_1 = R_{11}^0 C_1 + R_{22}^0 C_2 + R_{33}^0 C_3 \qquad (11\text{-}58)$$

where $R_{11}^0$, $R_{22}^0$, $R_{33}^0$ are the zero-frequency resistances seen by $C_1$, $C_2$, and $C_3$, respectively. Note that at zero frequency (dc), the capacitances are open-circuited and the $RC$ products in Eq. (11-58) are often referred to as *open-circuit time constants*.

We can give the following heuristic argument to justify the form of $a_1$ in Eq. (11-58). Consider $C_2 = C_3 = 0$ (open circuits) so that the circuit contains only $C_1$ as shown in Fig. 11-23*b*. This transfer function of the circuit in Fig. 11-23*b* has only one pole whose angular frequency is simply the reciprocal of the circuit time constant $\tau$. But $\tau$ for this case is $C_1$ multiplied by the equivalent resistance across its terminals, that is, $R_{11}^0$ as indicated in Fig. 11-23*b*. Note that letting $C_2 = C_3 = 0$ in Eq. (11-58) gives the same result. Similar arguments apply if we consider $C_1 = C_3 = 0$ and $C_1 = C_2 = 0$. For these conditions the time constants are $R_{22}^0 C_2$ and $R_{33}^0 C_3$, respectively.

[1] A formal proof of this method is given in Refs. 3 and 4 at the end of the chapter.

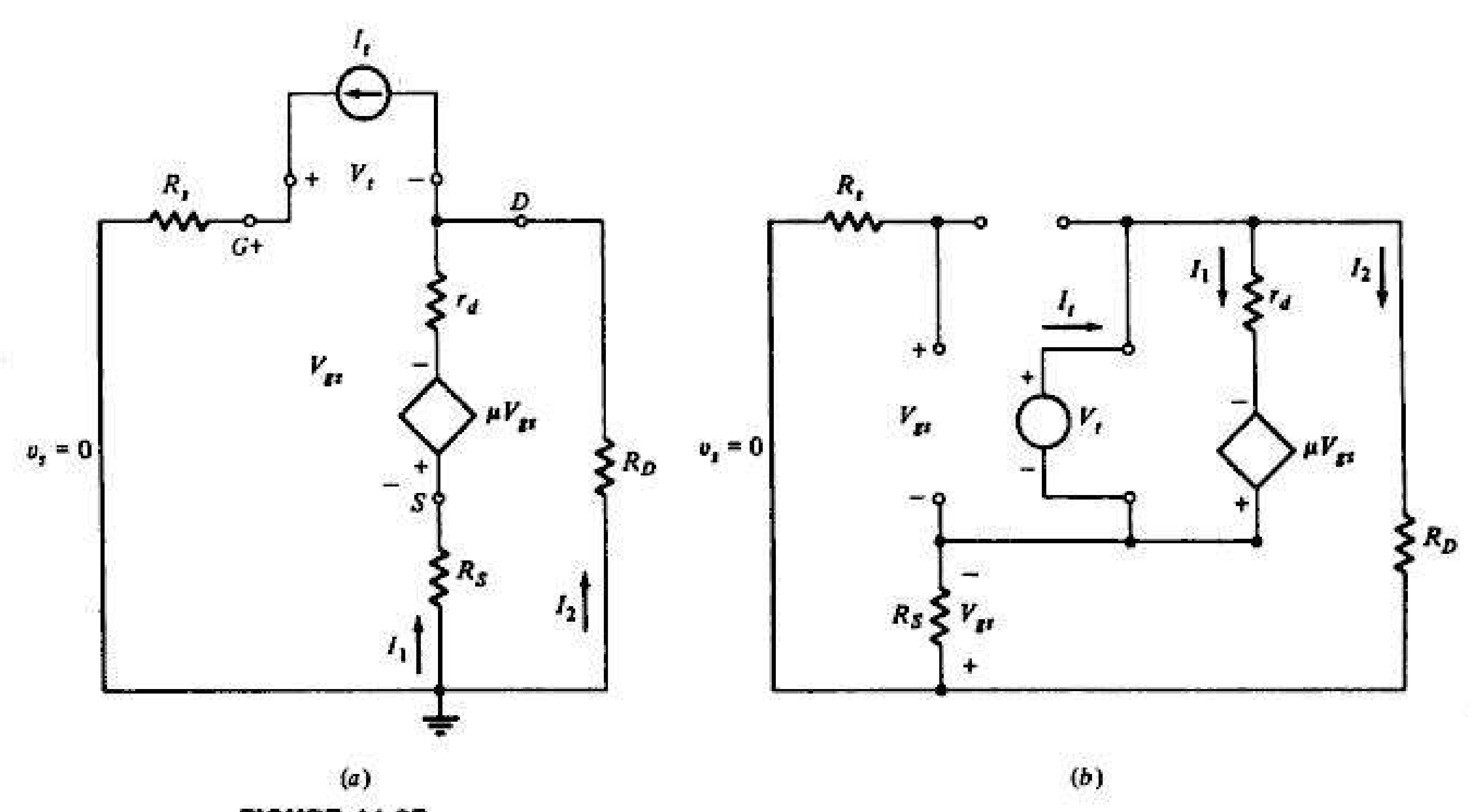



**FIGURE 11-27**
Circuit used to evaluate (a) $R^0_{gd}$ and (b) $R^0_{ds}$ for the stage in Fig. 11-25.

The control voltage $V_{gs} = I_2 R_S$. Combining these equations and solving for $I_t$, we obtain

$$R^0_{ds} = \frac{V_t}{I_t} = \frac{r_d(R_S + R_D)}{R_D + r_d + (1 + \mu) R_S}$$

(b) Numerical evaluation of $a_1$ is as follows:

$$\mu = g_m r_d = 2 \times 80 = 160$$

$$R_A = R_S \parallel (R_D + r_d) = 2 \parallel (20 + 80) = 1.96 \text{ k}\Omega$$

$$\mu' = \frac{\mu R_A}{r_d + R_D} = \frac{160 \times 1.96}{80 + 20} = 3.14$$

and

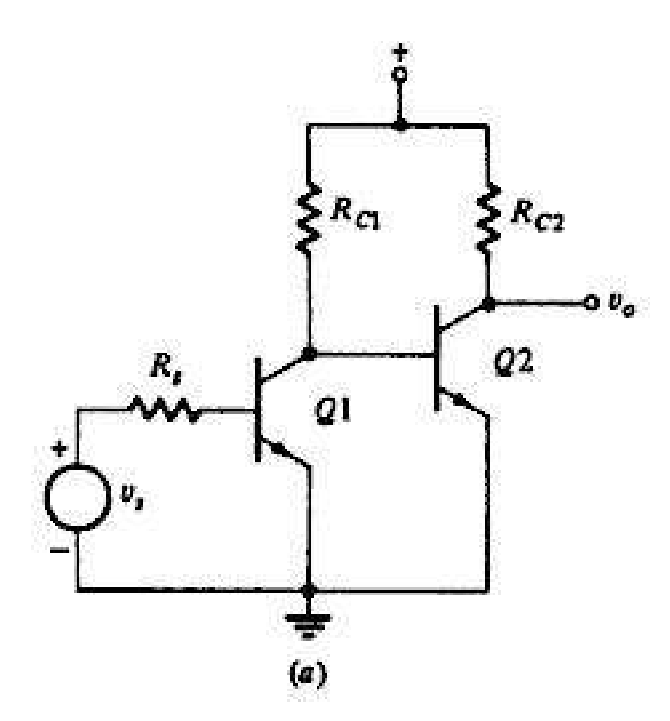
$$R^0_{gs} = \frac{0.3 + 1.96}{1 + 3.14} = 0.546 \text{ k}\Omega$$

$$R^0_{gd} = 0.3 + \frac{20[80 + 160 \times 0.3 + (1 + 160)2]}{20 + 80 + 2(1 + 160)} = 21.6 \text{ k}\Omega$$

$$R^0_{ds} = \frac{80(2 + 20)}{20 + 80 + (1 + 160)2} = 4.17 \text{ k}\Omega$$

Thus

$$a_1 = 0.546 \times 3 + 21.6 \times 1 + 4.17 \times 1.5 = 29.5 \text{ ns}$$



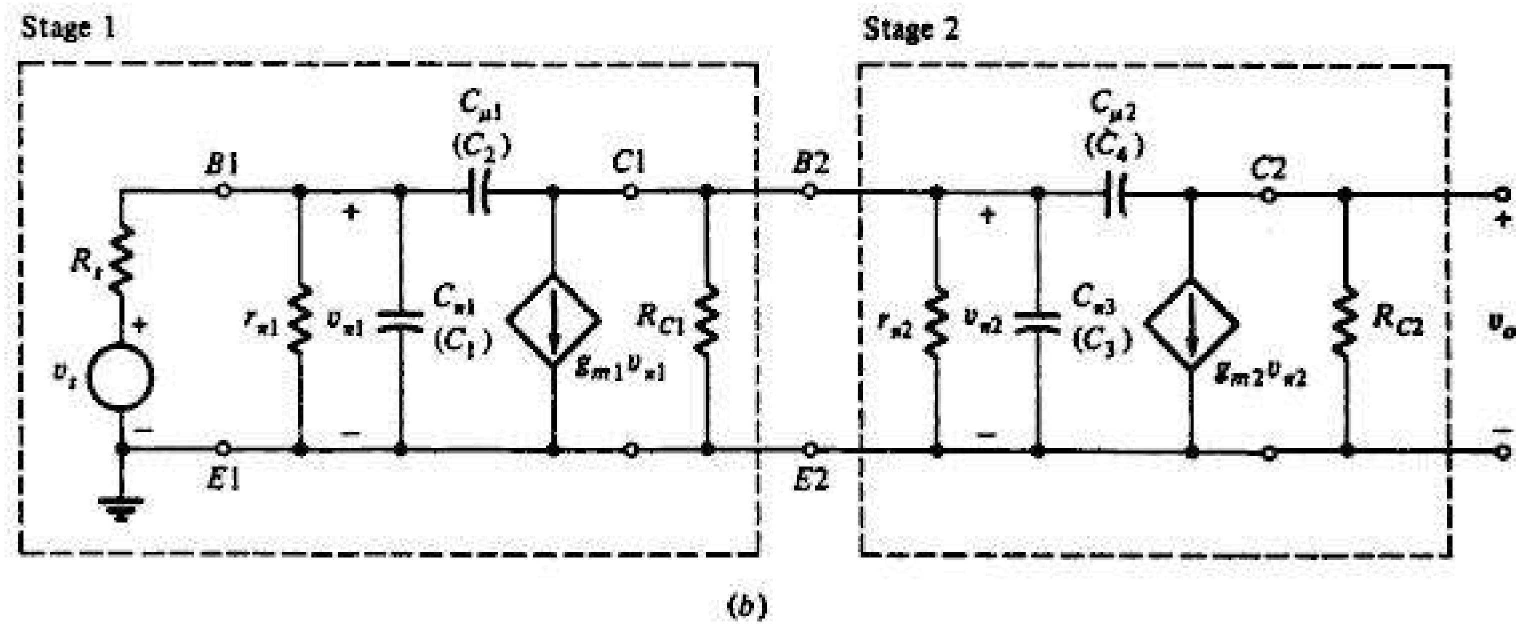




**FIGURE 11-30**
(*a*) A *CE-CE* cascaded amplifier and (*b*) its equivalent circuit at high frequencies.

relate the performance of individual stages to the overall response. Instead, we make use of our previous discussion of the single-stage amplifier to approximate the high-frequency response. On the basis of this analysis we can reasonably assume that the zeros of the transfer function occur at sufficiently high frequencies that they can be neglected. As we are primarily interested in evaluating the upper 3-dB frequency $f_H$, we assume that dominant-pole conditions apply, and all that is needed is to compute the coefficient $a_1$. This approach is justifiable because many practical amplifiers are designed to have a dominant pole. Furthermore, as we show in the following development, the $a_1$ coefficient of the cascade is readily related to the $a_1$ coefficients of the individual stages, a factor essential in amplifier design. On the basis of the evaluation of $a_1$ we can subsequently determine the coefficient $a_2$ (if necessary) and examine the validity of the dominant-pole approximation. We make this calculation in Example 11-5.

The circuit in Fig. 11-31 is used to evaluate the open-circuit resistances needed in the expression for $a_1$ [Eq. (11-59)]. To prevent the notation from becoming too cumbersome, it is convenient to identify $C_1 = C_{\pi1}$, $C_2 = C_{\mu1}$, $C_3 = C_{\pi3}$, and $C_4 = C_{\mu2}$ as indicated in Fig. 11-30*b*. We note that one advantage of open-circuiting the capacitances in Fig. 11-31 is that the stages are decoupled. Thus, from the results obtained in the analysis of the single-stage common-

The resistances $R_{33}^2$ and $R_{44}^2$ are

$$R_{33}^2 = 2.4 \parallel 1.5 \parallel 0.01 \parallel 0.4 = 0.01 \text{ k}\Omega$$
$$R_{44}^2 = 0.01(1 + 50 \times 0.6) + 0.6 = 0.91 \text{ k}\Omega$$

Use of these and previously determined values in part *a* yields

$$a_2 = 0.40 \times 24.5(0.923 \times 0.5 + 32.6) + 0.923 \times 19.5 \times 0.6 \times 0.5 + 38.2 \times 0.5(0.010 \times 19.5 + 0.91 \times 0.5) = 342 \times 10^{-18} \text{ s}$$

The angular frequency of the pole is

$$p_2 = \frac{a_1}{a_2} = \frac{61.5 \times 10^{-9}}{342 \times 10^{-18}} = 1.80 \times 10^8 \text{ rad/s}$$

and

$$f_2 = \frac{p_2}{2\pi} = \frac{1.80 \times 10^8}{2\pi} = 28.6 \text{ MHz}$$

As the poles are separated by more than a factor of 10, we can conclude that the dominant-pole approximation is valid. This is verified by computer simulation which gives $p_1 = 1.79 \times 10^7$ rad/s, $p_2 = 16.1 \times 10^7$ rad/s, and corresponding values $f_1 = 2.85$ MHz and $f_2 = 25.6$ MHz. The error in the dominant-pole calculation is nearly 10 percent, and the exact value of $f_H$ is 2.71 MHz. Thus $f_H = 1/2\pi a_1$ gives a very close approximation of the actual value. Note also that $f_H = 1/2\pi a_1$ is less than the actual value.

One further observation is noteworthy. Stages 1 and 2, performing as single-stage amplifiers, would have pole frequencies of 5.51 and 4.88 MHz, respectively. As a result of the coupling between the stages when they are cascaded, however, the two corresponding pole frequencies of the cascade are approximately 2.59 and 28.6 MHz. Hence one pole has moved closer to the origin and the other, further away. This condition is often referred to as *pole splitting* and is often exploited in compensating operational amplifiers.

## 11-11 THE CASCODE (*CE-CB*) AMPLIFIER

The cascode amplifier (Fig. 10-35) comprises a common-emitter stage in cascade with a common-base stage. In Sec. 10-17 we demonstrated that the midband gain of this combination is virtually the same as the gain of a common-emitter stage having the same load resistance $R_{C2}$ as does the common-base stage. Here we show that the frequency response of the composite is greater than that obtained for the corresponding common-emitter stage. To do so, we evaluate the $a_1$ coefficient of the cascode amplifier and compare the result with the $a_1$ coefficient of a common-emitter stage having a load $R_{C2}$.

The high-frequency equivalent circuit of the cascode amplifier is shown in Fig. 11-33*a*, and that used to evaluate the open-circuit resistances is given in Fig. 11-33*b*. Following the procedure outlined earlier in this section, the coefficient $a_1$ of the *CE-CB* cascade is the sum of the $a_1$ coefficients of the loaded common-emitter stage $a_{11}$ and the common-base stage $a_{12}$. For the common-emitter stage

mated by simply summing the lower 3-dB frequencies attributed to each capacitor in the circuit.

***Example 11-7*** A transistor amplifier stage has $R_E = R_C = 1.5\ \text{k}\Omega$, $R_s = 600\ \Omega$, and transistor parameters $\beta_o = 100$ and $r_\pi = 1.0\ \text{k}\Omega$. (*a*) Determine the values of $C_B$ and $C_E$ needed to obtain $f_L = 50$ Hz. Assume that each capacitor contributes equally to $f_L$. (*b*) Use the result in part *a* to determine the zero introduced by $C_E$.

***Solution*** (*a*) For an overall $f_L = 50$ Hz, $f_{LB} = f_{LE} = 25$ Hz. Using Eqs. (11-76) and (11-80), we obtain

$$f_{LB} = 25 = \frac{1}{2\pi(0.6 + 1.0)C_B} \qquad \text{or} \qquad C_B = 3.98\ \mu\text{F}$$

$$f_{LE} = 25$$

$$= \frac{1}{2\pi C_E\{[1.5(0.6 + 1.0)/(1 + 100)]/[1.5 + (0.6 + 1.0)/(1 + 100)]\}}$$

or $\quad C_E = 406\ \mu\text{F}$

(*b*) From Eq. (11-66), $f_E = \omega_E/2\pi = 1/2\pi R_E C_E = 1/(2\pi \times 1.5 \times 0.406) = 0.261$ Hz.

The capacitance values obtained in Example 11-7 indicate the typical situation that exists in practice, namely, that the value of $C_E$ required is much larger than the value of $C_B$. In addition, the zero introduced by $C_E$ occurs at such a low frequency compared with the desired value of $f_L$ that it has negligible effect on the low-frequency response. Both size and cost of discrete capacitances increase with increasing capacitance value. Consequently, in amplifier design, it is customary practice to select $C_E$ to meet the specified value of $f_L$. Then $C_B$ is selected to make $f_{LB}$ occur at a much lower frequency than $f_L$. A good rule of thumb is to choose $f_{LB} \le f_L/10$. For the values in Example 11-7, this gives $C_E = 203\ \mu\text{F}$ and $C_B \ge 19.9\ \mu\text{F}$, thus making the total capacitance about one-half of that found in Example 11-7.

**Cascaded Stages at Low Frequencies** The lower 3-dB frequency $f_L$ of cascaded amplifiers is readily obtained by extension of the method described in Sec. 11-13 for the single-stage amplifier. Originally given as Eq. (11-85) and repeated as Eq. (11-86) in a somewhat altered form, $f_L$ is expressible as

$$f_L = \sum_{K=1}^{N} f_{LK} = \frac{1}{2\pi}\sum_{K=1}^{N}\frac{1}{R_{KK}^{z}C_K} \tag{11-86}$$

where $R_{KK}^{z}C_K$ is the circuit time constant when all other capacitors are short-circuited. It is clear in Eq. (11-86) that the overall lower 3-dB frequency $f_L$ is related to the lower 3-dB frequencies of the individual stages. Note that the rule of thumb given immediately after Example 11-7 also applies to cascaded stages at low frequencies.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

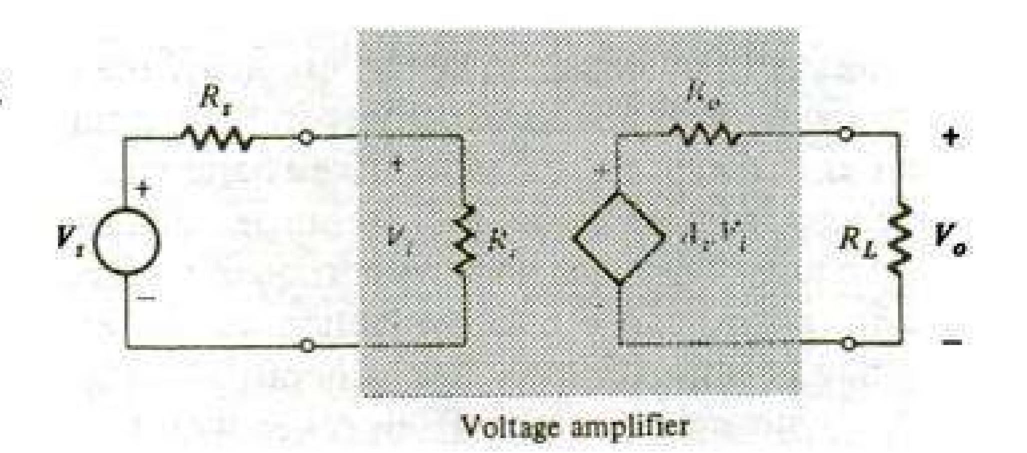



**FIGURE 12-1** Equivalent circuit of a voltage amplifier.

input resistance is much greater then $R_s$, then $V_i \approx V_s$. Similarly, if $R_L \gg |R_o|$, then $V_o \approx A_vV_i = A_vV_s$. This amplifier provides an output voltage proportional to the input voltage, and *the proportionality factor is independent of the magnitudes of the source and lead resistances.* Such a circuit is called a *voltage amplifier* or a *voltage-voltage converter* and its behavior is that of the voltage-controlled voltage source. The symbol $A_v$ in Fig. 12-1 represents $V_o/V_i$ with $R_L \to \infty$ and hence is the *open-circuit voltage gain.*

**The Current Amplifier** An ideal current-controlled current source is unilateral, has zero input resistance $R_i$, and has infinite output resistance $R_o$. The practical *current amplifier*, or *current-current converter*, shown in Fig. 12-2 approaches ideal behavior when $|R_i| \ll R_s$, so that $I_i \approx I_s$ and when $R_L \ll |R_o|$, making $I_o = A_iI_i \approx A_iI_s$. Thus the output current is proportional to the input current and independent of either $R_s$ or $R_L$. Note that $A_i \equiv I_o/I_i$, with $R_L = 0$, represents the *short-circuit current gain*. The circuit in Fig. 12-2 is analogous to the simplest BJT model based on the Ebers-Moll equations in which $I_c = \beta I_b$.

**The Voltage-Current Converter or Transconductance Amplifier** The *voltage-current converter* or *transconductance amplifier* is based on the ideal voltage-controlled current source. Note the similarity of this amplifier type, shown in Fig. 12-3, to the unilateral hybrid-$\pi$ model of the BJT. To approximate ideal characteristics, $R_s \ll |R_i|$ and $R_L \ll |R_o|$ in the practical voltage-current converter. These conditions make $V_i \approx V_s$ and $I_o \approx G_mV_i \approx G_mV_s$, so that $G_m$, the proportionality factor, is independent of both load and source resistances. The parameter $G_m \equiv I_o/V_s$ with $R_L = 0$ is referred to as the *short-circuit transfer conductance* (or simply the *transconductance*). Note the similarity of $G_m$ for the overall amplifier with the definition of $g_m$ for the transistor.

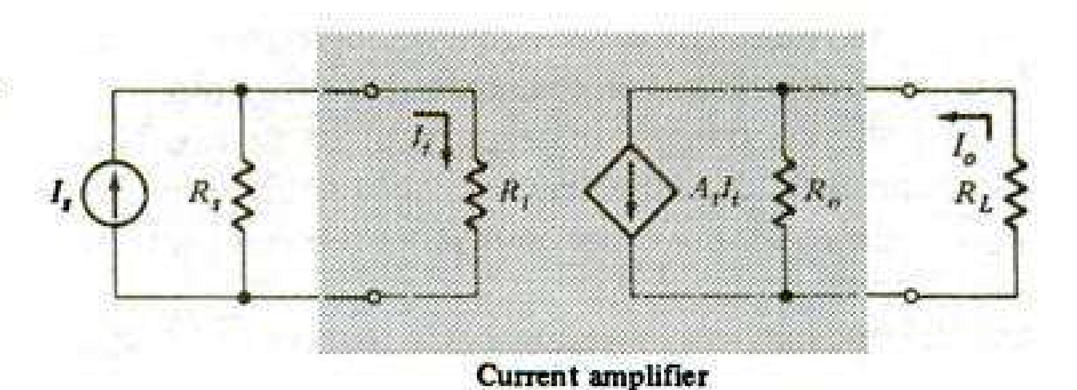



**FIGURE 12-2** Current amplifier equivalent circuit.

connection is depicted in Fig. 12-7*b*, in which source current $I_s$ and feedback current $I_f$ are compared. Note that the amplifier input current $I_i$ is proportional to the difference $I_s - I_f$.

**The Feedback Network** This block in Fig. 12-5 is usually a passive network which may contain resistors, capacitors, and inductors. Most often it is a purely resistive network. In several classes of integrated circuits (ICs) discussed in Part 4, capacitors or resistor-capacitor combinations form the feedback network. One function of the feedback network is to convert the sample of the output signal to the form of the signal appropriate for comparison. For example, consider a feedback amplifier in which the output is a voltage and currents are compared at the input. The transmission from the output port to the input port of the feedback network must convert the output voltage to a proportionally valued current at the input.

**The Basic Amplifier** The basic amplifier in Fig. 12-5 is one of the four configurations given in Figs. 12-1 to 12-4. This circuit amplifies the difference signal that results from comparison, and it is this process that is responsible for the desensitivity and control of the output in a feedback system.

Consider the feedback amplifier in Fig. 12-8, in which the basic amplifier is the current amplifier in Fig. 12-2. Let us assume that $A_i$ increases as a result, perhaps, to an increase in $\beta_o$, in one of the transistors which comprise $A_i$. An increase in $A_i$ tends to increase the load current $I_o$, and hence the feedback current $I_f$ also increases. Neglecting the current in $R_s$, the control current $I_i = I_s - I_f$ decreases. With reduced input drive, the amplifier output tends to decrease and thus offsets the effect of the increased $A_i$. This action is the basis of *negative feedback*. Because the summing network provides a difference signal, the amplifier input changes in a direction opposite to the change in the output. The net effect is to maintain a constant output signal that is independent

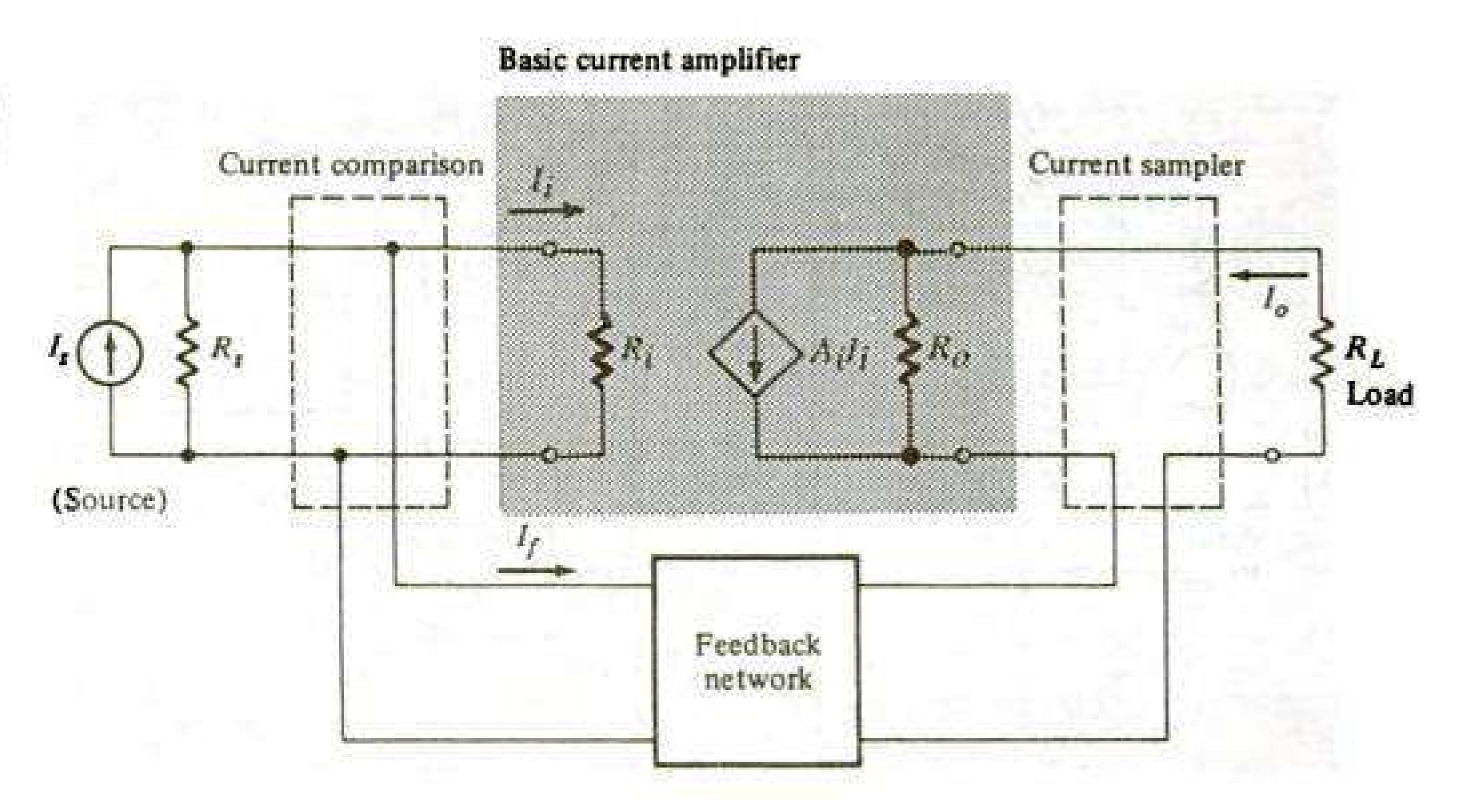



**FIGURE 12-8**
Feedback amplifier with current summing and current sampling.

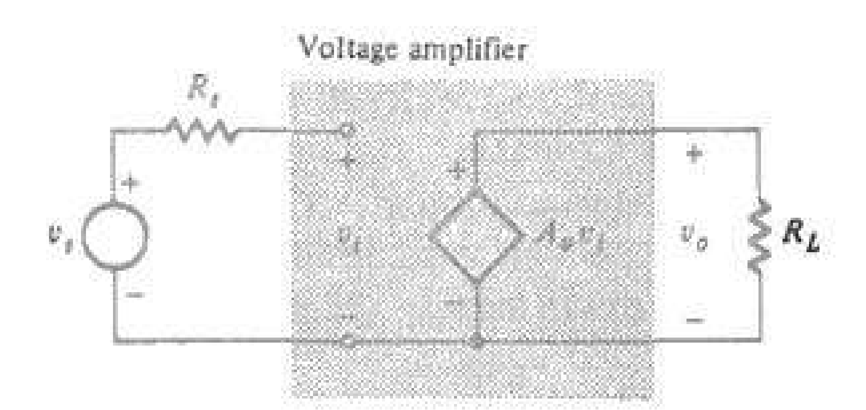



**FIGURE 12-10**
Ideal voltage amplifier.

Since $\beta$ is the transfer function of the usually passive feedback network, $A_F$ is essentially *independent of the gain of the basic amplifier and depends only on the ratio of passive components*. This is the situation we encountered in the inverting and noninverting Op-Amp stages (Sec. 10-21) for which the gain of these stages was proportional to the resistance ratio $R_2/R_1$.

Although it is exact only for small variations in $T$, Eq. (12-11) provides an estimate of the improvement in sensitivity; that is, the percentage change in $A_F$ is the percentage change in $T$ divided by $(1 + T)$. For the numerical values given above, the percentage change in $A_F$ is 1 percent for 50 percent changes in $T$. Note that while Eq. (12-11) indicates that $\Delta A_F/A_F$ is the same for positive and negative changes in $T$, this is not the case for large variations in $T$.

**Nonlinear Distortion** The analysis in this and Chaps. 10 and 11 assumed that the amplifier stages behaved linearly; that is small-signal conditions applied. However, if a large signal is applied, the amplifier characteristic exhibits nonlinearity and the output waveform becomes distorted (Fig. 10-2*b*). We illustrate the effect of feedback on nonlinear distortion in the following discussion.

The voltage amplifier in Fig. 12-10 is ideal in all respects except that its dynamic range is limited. That is, the range of input-signal amplitudes that can be accommodated for linear operation is restricted. This is depicted in the voltage transfer characteristic in Fig. 12-11, which displays the relationship

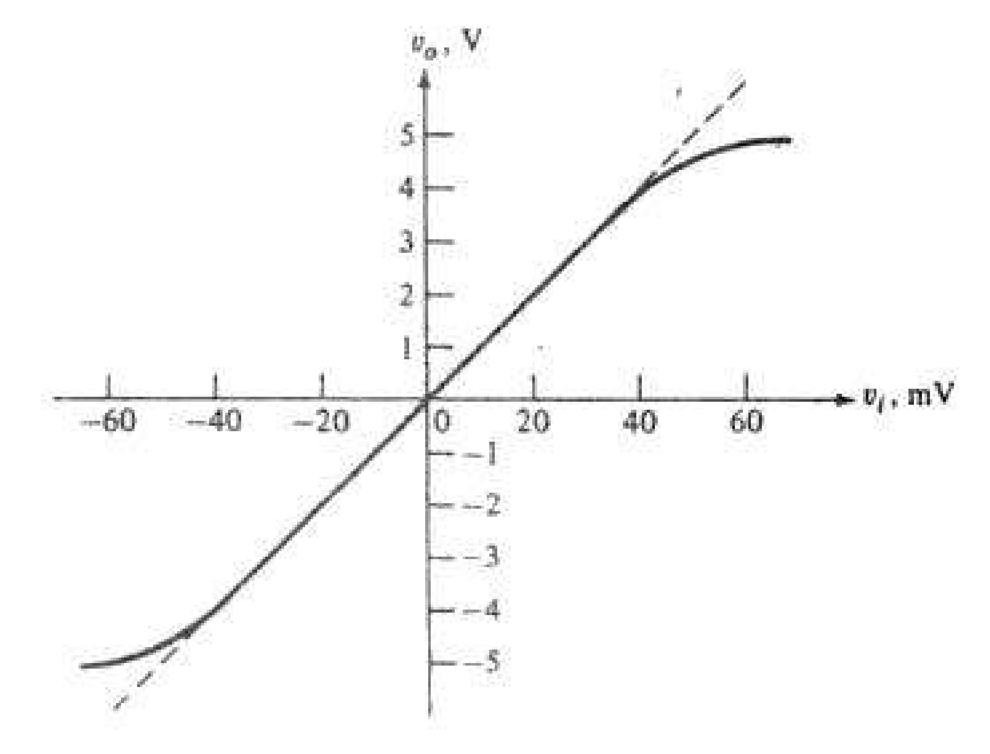



**FIGURE 12-11**
Voltage transfer characteristic for amplifier in Fig. 12-10.

feedback. If $1 + T$ is much larger than unity, this would seem to represent a considerable reduction in the output noise. However, as noted above, for a given output the amplification of the preamplifier for a specified overall gain must be increased by the factor $1 + T$. Since the noise generated is independent of the signal amplitude, there may be as much noise generated in the preamplifying stage as in the output stage. Furthermore, this additional noise will be amplified, as well as the signal, by the feedback amplifier, so that the complete system may actually be noisier than the original amplifier without feedback. Special, low-noise preamplifiers are used in several applications, such as high-quality stereo systems, to exploit the benefits of feedback and improve the signal-to-noise rates. If the additional gain required to compensate what is lost because of the presence of negative feedback can be obtained by a readjustment of the circuit parameters rather than by the addition of an extra stage, a definite reduction will result from the presence of the feedback. In particular, the hum introduced into the circuit by a poorly filtered power supply may be decreased appreciably.

## 12-5 IMPEDANCE IN FEEDBACK AMPLIFIERS

Earlier in this chapter we stated that feedback is used to make practical amplifier characteristics approximate those of ideal amplifiers. To do so requires that the input and output resistances (impedances) of the feedback amplifier have appropriate values (Table 12-1). We now examine the effect of the topology of a feedback on the amplifier on these impedance levels.

### Input Resistance

*If the feedback signal is returned to the input in series with the applied voltage, the input impedance increases.*[1] The preceding is valid independent of the output connection. Thus the increase in input resistance is exhibited in both the series-shunt and series-series configurations.

Consider the circuit in Fig. 12-15$a$, which depicts the series-connected input circuit of a feedback amplifier. The KVL for the loop gives

$$V_i = V_s + V_f$$

The feedback signal is $V_f = \beta X_o$, where $X_o$ is the output signal and $X_o = AV_i$. Combination of these relationships gives

$$V_i = IR_i = \frac{V_s}{1 - A\beta}$$

from which the resistance with feedback $R_{IF}$ is

$$R_{IF} \equiv \frac{V_s}{I} = R_i(1 - A\beta) = R_i(1 + T) \tag{12-15}$$

[1] Although we treat circuits only at midband frequencies in this chapter, the relationships developed apply at all signal frequencies.

In Eq. (12-21), we have the following conditions:

1. $Z_F$ is the impedance seen looking into any pair of terminals $A$ and $B$ of a feedback amplifier
2. $Z_D$ is the dead-system value of $Z_F$, that is, it is the impedance seen looking in to terminals $A$ and $B$ when the gain of the amplifier is made zero.
3. $T_{SC}$ is the return ratio measured when terminals $A$ and $B$ are short-circuited.
4. $T_{OC}$ is the return ratio measured when terminals $A$ and $B$ are open-circuited.

Note that Eq. (12-21) reduces to Eq. (12-19) if $T_{OC} = 0$ and to Eq. (12-20) when $T_{SC} = 0$.

Blackman's impedance formula applies to all feedback amplifiers and not simply to the ideal situation described in Figs. 12-15 and 12-16. Embodied in the quantities $T_{OC}$, $T_{SC}$, and $Z_D$ are the loading effects of the feedback network $\beta$ on the basic amplifier $A$ and the effects of the load and source resistances $R_s$ and $R_L$, respectively, on the values of $A$ and $\beta$. In the analysis of the four basic single-loop topologies in the next several sections, we include these loading effects.

## 12-6 PROPERTIES OF FEEDBACK AMPLIFIER TOPOLOGIES

Some of the general characteristics of single-loop feedback amplifiers were described in the four previous sections. Each of the four topologies introduced in Sec. 12-2 approximates one of the four amplifier types (Sec. 12-1). In this section, we investigate specific characteristics of the four topologies. Transistor realizations that approximate these circuits are discussed in the remainder of the chapter.

### The Shunt-Shunt Amplifier

The two-port network representation of a shunt-shunt amplifier is shown in Fig. 12-17. The parallel connection at the output signifies that the output voltage is sampled, whereas the parallel-connected input results in current comparison. Thus the feedback network provides voltage-to-current transfer.

The use of Thévenin or Norton's theorem allows representation of the internal amplifier by any of the four amplifier types discussed in Sec. 12-1. Note that $V_i = z_i I_i$ and that the current source $gV_i$ in parallel with $r_o$ can be converted to its voltage-source equivalent. The representation in Fig. 12-17 is based on the two-port $y$ parameters.[1]

We expect networks in parallel to have low impedance levels; this is demonstrated by using Blackman's impedance formula [Eq. (12-21)] as follows. When terminals 1–1′ in Fig. 12-17 are short-circuited, $I_i$ and $V_i$ are zero, and

[1] Two 2-port networks in parallel can be represented by an equivalent network having $y$ parameters equal to the sum of the $y$ parameters of the constituent networks. In practical circuits, however, it is often difficult to identify the individual networks.

$$R_{OF} = \frac{r_o}{1 + [\beta_o r_o/(R_s + r_\pi)]} = \frac{r_o(R_s + r_\pi)/\beta_o}{r_o + (R_s + r_\pi)/\beta_o} \approx \frac{R_s + r_\pi}{\beta_o}$$

for $r_o \gg (R_s + r_\pi)/\beta_o$. For $\beta_o \gg 1$, $R_{OF}$ is the entry given for the output resistance in Table 10-3*A*.

***Example 12-2*** Determine $A_F$, $T$, and $R_{OF}$ for the common-source stage with source resistance in Fig. 12-22*a*.

***Solution*** The input circuit is analogous to that for the emitter follower and is thus series-connected. Making $V_o = 0$ does not eliminate the feedback because $I_o$ and, hence, $V_f$ do not become zero. When $I_o = 0$, $V_f = 0$ and the output is also series connected. (This amplifier is a series-series type.)

To determine the input circuit of the amplifier without feedback, open-circuit the output ($I_o = 0$). The feedback resistance $R_S$ appears in series with $V_s$ as shown in Fig. 12-22*b*. To find the output circuit, set $I_i = 0$. Again, as indicated in Fig. 12-22*b*, $R_S$ appears in the output loop. The equivalent circuit in Fig. 12-22*b* is depicted in Fig. 12-22*c*.

In Fig. 12-22*c*, $V_o = -I_o R_D$ and $V_f = -I_o R_S$; hence $\beta = V_f/I_o = -R_S$. Since no current exists in the gate loop, $V_{gs} = V_s$. Use of KVL for the drain loop gives $I_o = \mu V_s/(r_d + R_D + R_S)$. It follows that

$$A_{OL} = \frac{I_o}{V_s} = \frac{\mu}{r_d + R_D + R_S}$$

The return ratio is

$$T = -\beta A_{OL} = \frac{\mu R_S}{r_d + R_D + R_S}$$

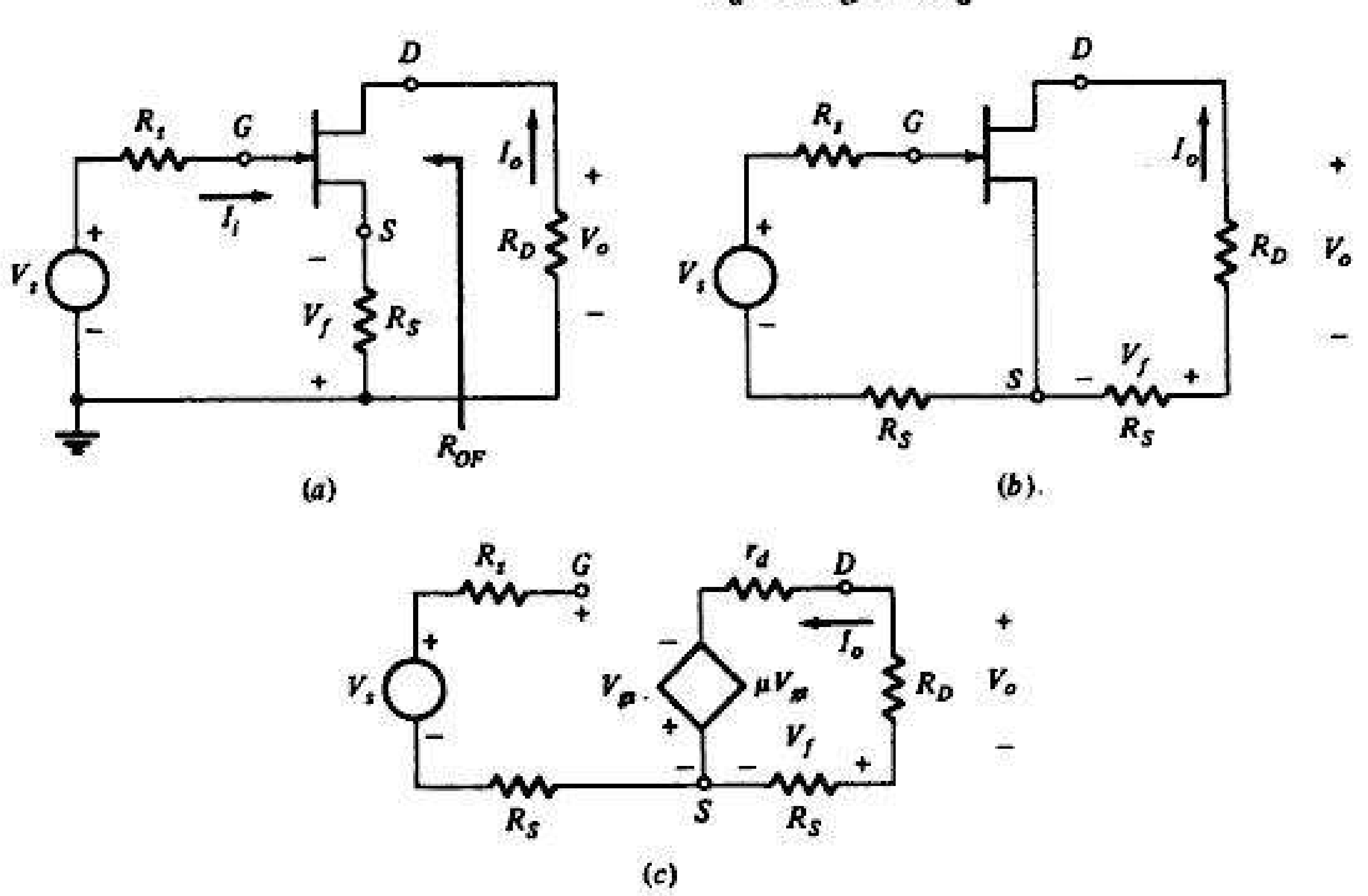



**FIGURE 12-22** (*a*) Common-source amplifier with source resistance. (*b*) Schematic diagram and (*c*) equivalent circuit of the amplifier without feedback.

The parameter $K$ depends only on the passive elements which comprise $t_{21}$ and $t_{22}$.

If $|A_D| \ll |KT|$, then $A_{OL} = KT$ and

$$A_F \approx \frac{KT}{1 + T} \tag{12-32}$$

For $T \gg 1$, $A_F \approx K = -1/\beta$ and illustrates that the closed-loop gain is essentially independent of the basic amplifier gain and depends only on the ratio of passive components. This is the situation we encountered in the Op-Amp stages of Sec. 10-21 for which the gain was proportional to the resistance ratio $R_2/R_1$.

Observe that Eqs. (12-32) and (12-9) are identical because in Eq. (12-32) we neglect $A_D$, the feedforward transmission.

We can give additional meaning to $K = -1/\beta$ [Eq. (12-7)] by examining Eq. (12-22). If the gain of the basic amplifier is infinite ($t_{12}$ and hence $A_{OL}$ and $T$ become infinite) and $X_o$ *remains finite*, the input to the amplifier $X_i = \hat{X}_i = 0$. Setting $X_i = 0$ in Eq. (12-22) indicates that $X_o/X_s = -(t_{21}/t_{22}) = K$ and is the gain of the feedback amplifier when $T \to \infty$. Making $X_i = 0$ is equivalent to stating that the feedback signal $t_{22}X_o$ and the component of the input signal $t_{21}X_s$ are equal in magnitude and opposite in phase. We have encountered this situation in Sec. 10-21 in the discussion of the inverting Op-Amp stage. There we showed that when $A_v \to \infty$, $V_i \to 0$ and the current produced by the signal source $V_s/R_1$ was balanced by the feedback signal $V_o/R_2$.

**The Analysis Procedure** For a specific circuit, the $t$ parameters are determined by first assuming the controlled source in the device model is an independent source; thus we must first identify $\hat{X}_i$ and then apply Eqs. (12-24). This is illustrated in the two examples that follow. We have deliberately selected two circuits previously analyzed so that we may focus on the techniques used to obtain the $t$ parameters and compare the results with prior analyses.

***Example 12-3*** (*a*) Evaluate the $t$ parameters for the emitter follower in Fig. 12-24*a*. (*b*) Use the results in part *a* to obtain $A_F$, $T$, $A_D$, and $K$.

***Solution*** (*a*) The equivalent circuit of the emitter follower is given in Fig. 12-24. The parameters $X_s = V_s$, $X_i = V_\pi$, $X_o = V_o$, and $\hat{X}_i = \hat{V}_\pi$ are identified in the equivalent circuit. Note that $\hat{V}_\pi$ is associated with the controlled source, thus making $g_m V_\pi$ behave as an independent current source.

To determine $t_{11}$ by use of Eq. (12-17), $\hat{X}_i = \hat{V}_\pi$ is suppressed and Fig. 12-24 reduces to the circuit in Fig. 12-25*a*. In Fig. 12-25*a*, use of the voltage-divider relationship gives

$$t_{11} = \left.\frac{V_o}{V_s}\right|_{\hat{V}_\pi = 0} = \frac{R_E}{R_s + r_\pi + R_E}$$

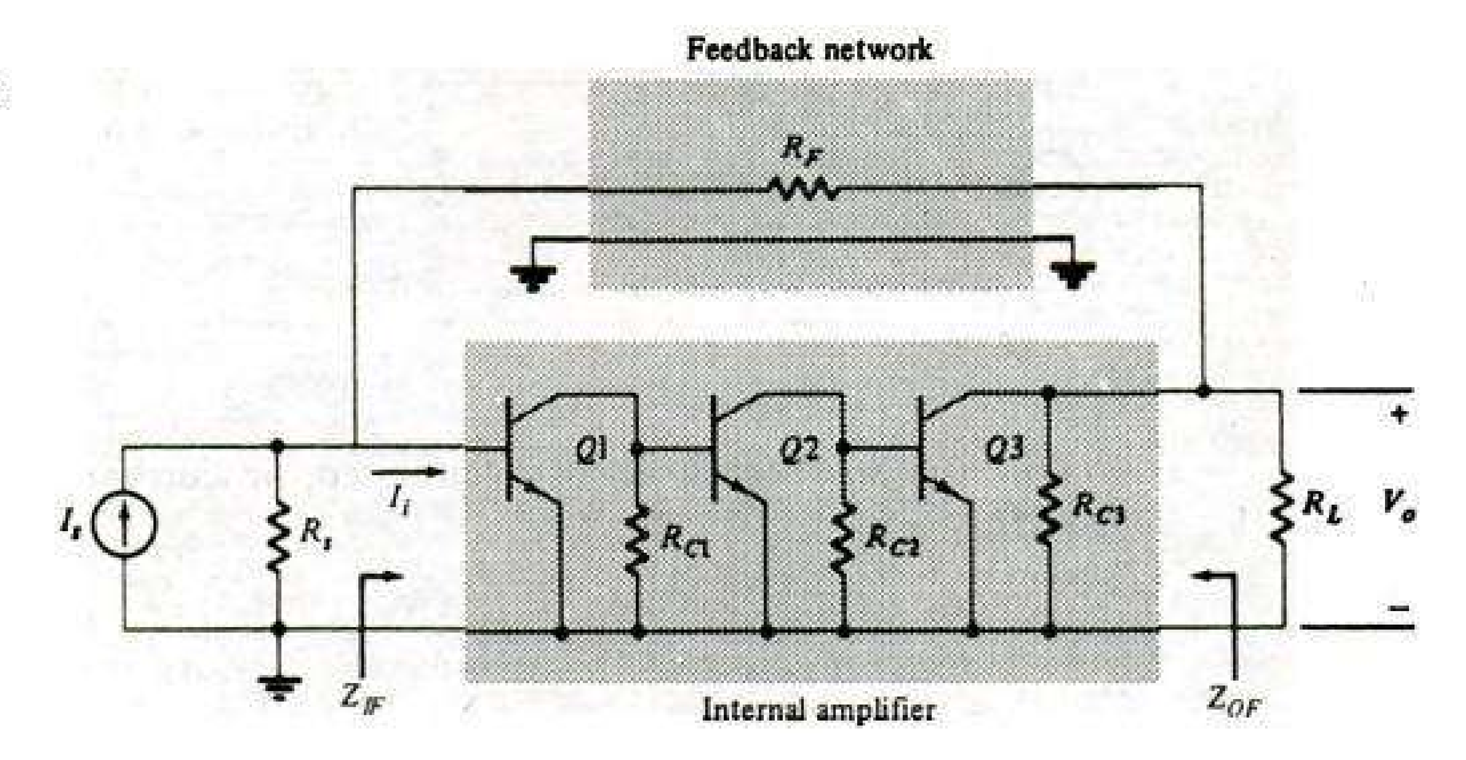



**FIGURE 12-29**
A bipolar shunt-triple feedback amplifier.

single-loop feedback topologies. For example, transistor stages have inherent feedback provided by $C_\mu$ ($C_{gd}$) and, therefore, are not truly single-loop amplifiers at high frequencies. Also, when common-emitter (common-source) stages containing $R_E$ ($R_S$) are used as one stage of the basic amplifier, even at low frequencies, they are feedback amplifiers. The type of feedback described in the two prior sentences is called *local feedback* because the feedback loop is around only one stage. In multistage circuits, we are concerned with *global feedback*, that is, the overall feedback around a number of stages in cascade.

The three-stage amplifier in Fig. 12-29 has the shunt-shunt structure of Fig. 12-17 and is commonly called a *shunt triple*. The internal three-stage amplifier identified in Fig. 12-29 can be modeled as a single equivalent amplifier as depicted in Fig. 12-30. The controlled source $Z_m I_i$ and series resistance $r_o$ are the Thévenin equivalent of the amplifier block in Fig. 12-29. The resistance $r_i$[1]

[1]Note that $r_b$ is included. Because the shunt connection has a very low input resistance, neglecting of $r_b$ *may* sometimes introduce an error.

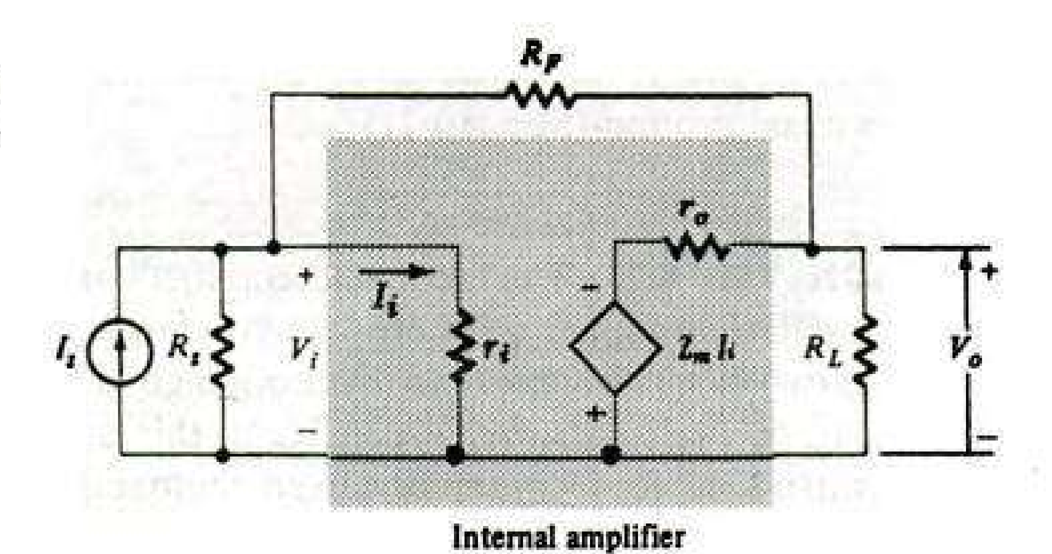



**FIGURE 12-30**
The equivalent-circuit representation of the shunt triple.

**12-11 THE SHUNT-SERIES PAIR** The two-stage amplifier in Fig. 12-32 is called the *shunt-series pair*. Clearly, current comparison takes place at the input. We now show that current sampling occurs at the output by demonstrating that the feedback current $I_f$ is proportional to the output current $I_o$. Assuming that $\beta_o \gg 1$, the emitter current in $Q2$ is $I_o$. The paths through the resistances $R_E$ and $R_F$ form a current divider and, hence $I_f$ is proportional to $I_o$. In addition, setting $I_o = 0$ causes $I_f$ to be zero.

The circuit in Fig. 12-32 only approximates the shunt-series amplifier of Fig. 12-19 because of the local feedback in $Q2$ provided by $R_E$.

The amplifier without feedback shown in Fig. 12-33 is used to determine the approximate value of $A_{OL} = I_o/I_s$. As seen in Fig. 12-33, the loading effect of the feedback network, obtained by using the rules in Sec. 12-7, is included. We also note that only the global feedback loop is eliminated; the local feedback due to the emitter resistance in $Q2$ is included as part of $A_{OL}$. Analysis of the circuit in Fig. 12-33 (Prob. 12-30), making the following assumptions listed, gives the results stated in Eqs. (12-48).

$$A_{OL} \approx \frac{-\beta_o R_{C1}}{R_E} \qquad T \approx \frac{\beta_o R_{C1}}{R_F} \qquad K = -\frac{1}{\beta} = \frac{-R_F}{R_E} \tag{12-48}$$

where it is assumed that $\beta_o \gg 1$, $\beta_o R_E \gg$ both $r_{\pi 2}$ and $R_{C1}$, $R_s \gg r_{\pi 1}$, and $R_F \gg$ both $R_E$ and $r_{\pi 1}$.

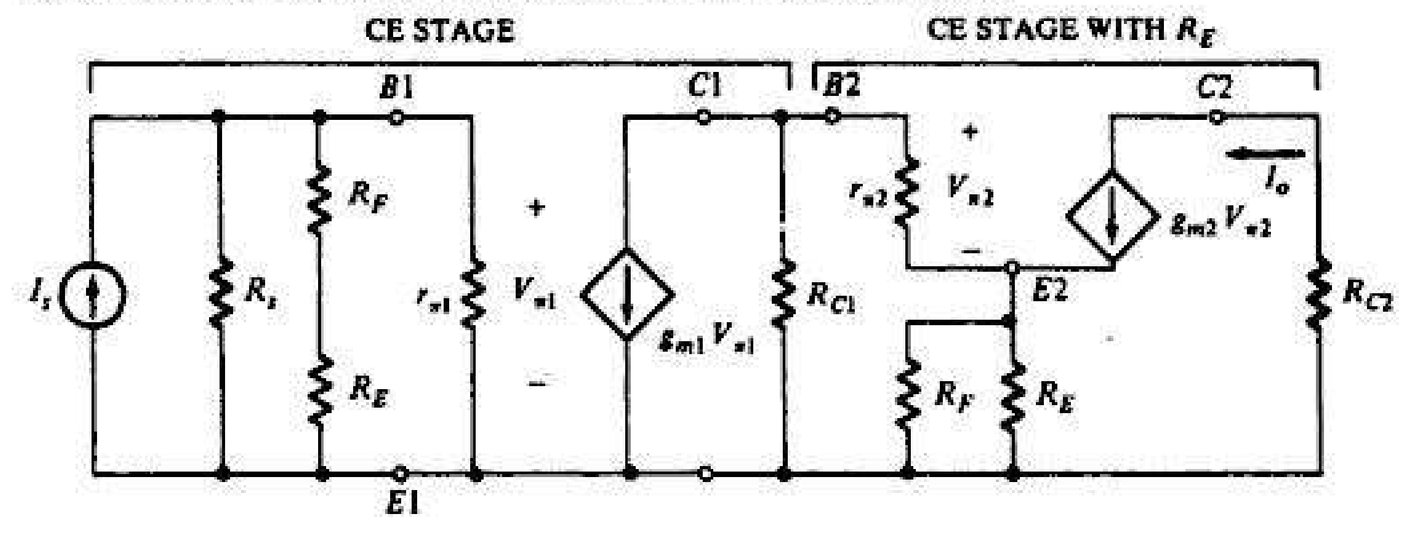
**The Input and Output Resistances** The input impedance $Z_{IF}$ is reduced because of the shunt connection at the input. Similar to the discussion in the previous section, we obtain

$$Z_{IF} = \frac{r_{\pi 1}}{1 + T\,|_{R_s \to \infty}} \tag{12-49}$$

where it is assumed that $R_F \gg r_{\pi 1}$.

If $r_{o2} \to \infty$, then the dead-system impedance is also infinite. When $r_{o2}$ is included in the model (Prob. 12-29) $Z_{OD}$, assuming $R_F \gg R_E$, is

**FIGURE 12-33**
The model of the basic amplifier of the shunt-series pair including the loading effect of the feedback network at both input ($R_F$ in series with $R_E$) and output ($R_F \| R_E$).



Note that the last term in Eq. (12-57) is the voltage-to-current transfer ratio of the third stage.

In Fig. 12-35*b* letting $V_{gs1} = 0$ results in $V_f = I_o R_S$. Hence $\beta = V_f/I_o = R_S$ and the return ratio can be approximated by

$$T = -A_{OL}\beta \tag{12-58}$$

$$= \frac{\mu_1 \mu_3 g_{m2} R_{D1} R_{L2} R_S}{[R_{D1} + r_{d1} + (1 + \mu_1)R_S]\,[R_{D3} + r_{d3} + (1 + \mu_3)R_F]}$$

**The Input and Output Impedances** The input resistance of this stage is virtually infinite at midband frequencies. The gate-to-source resistance of the MOSFET is in the order of $10^{12}\ \Omega$, and this value is multiplied by $1 + T_{SC}$ because of the series feedback. Consequently, $Z_{IF}$ for this circuit is usually considered as an open circuit. In discrete-transistor stages, the input impedance is simply the equivalent resistance of the gate-bias network.

The dead-system output impedance $Z_{OD}$ is the output resistance of a common-source stage containing a source resistance as determined in Table 10-5. Assuming $R_F \gg R_S$, we have

$$Z_{OF} = [r_{d3} + (1 + \mu_3)R_F]\,(1 + T\,|_{R_{D3}=0}) \tag{12-61}$$

where $T_{SC}$ is obtained from $T$ by setting $R_{D3} = 0$.

*Example 12-9* The FET parameters and drain resistances used in the series triple in Fig. 12-35 are listed in Table 12-6. The feedback resistor $R_F = 10\ \text{k}\Omega$, and $R_S = 0.50\ \text{k}\Omega$. Find $A_{OL}$, $T$, and $A_F$ for this stage.

**TABLE 12-6 FET Parameters and Component Values for Example 12-9 and Fig. 12-35**

| | *Parameter* | | | |
|---|---|---|---|---|
| *Transistor* | $g_m$, m℧ | $r_d$, kΩ | $\mu$ | $R_D$, kΩ |
| *Q*1 | 3.0 | 50 | 150 | 50 |
| *Q*2 | 1.5 | 100 | 150 | 50 |
| *Q*3 | 1.0 | 130 | 130 | 10 |

*Solution* We use the results of the approximate analysis in Eqs. (12-57) and (12-58) to evaluate $A_{OL}$ and $T$, respectively. Note that the assumption that $R_F \gg R_s$ is reasonable and will introduce only a small error. In Eq. (12-57), the value of $R_{L2}$ used is

$$R_{L2} = R_{D2} \parallel r_{d2} = 50 \parallel 100 = 33.3\ \text{k}\Omega$$

Then

$$A_{OL} = \frac{-150 \times 50}{50 + 50 + (1 + 150)\,0.5}\,(-1.5 \times 33.3) \times \frac{130}{10 + 130 + (1 + 130)\,10} = 191.6\ \text{m℧}$$

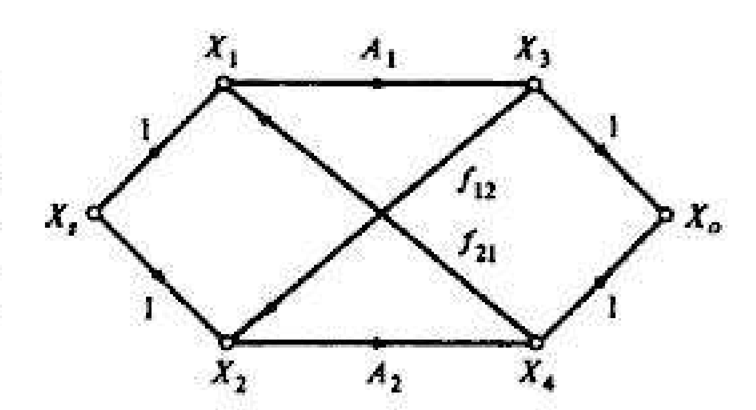



**FIGURE 12-40**
A two-channel McMillan feedback-feedforward amplifier. Only one channel must operate normally to provide the desired output.

Nominally, the two channels are designed to be identical with $A_1 = A_2 = A$ and $f_{12} = f_{21} = f$. Under these circumstances

$$A_F = \frac{2A\,(1 - Af)}{1 - A^2f^2} = \frac{2A\,(1 - Af)}{(1 + Af)\,(1 - Af)} = \frac{2A}{1 + Af} \tag{12-62}$$

If $Af = 1$, then, from Eq. (12-62), $A_F = A$.

Suppose that $A_1 = 0$, that is, a catastrophic failure occurs in one channel. Use of Eq. (12-61) indicates that $A_F = A_2 = A$ and the same transmission between input and output exists. Recall that $A_2$ is a negative-feedback amplifier, so that even with $A_1 = 0$, the circuit performance displays the benefits of a single-loop amplifier.

The amplifier structure in Fig. 12-40 is also insensitive to any changes that occur in one channel. Thus, if $A$ changes so that $A_1f_{12} \neq 1$ while $A_2f_{21} = 1$, from Eq. (12-61) we have

$$A_F = \frac{A_2\,(1 - A_1f)}{1 - A_1f} = A_2 = A$$

and no variation in $A_F$ exists. Similarly, if $f_{12}$ changes so that $Af_{12} \neq 1$, then

$$A_F = \frac{A\,(1 - Af_{12})}{1 - Af_{12}} = A$$

We conclude that this amplifier is invariant to changes in both active and passive elements. Again, one drawback is that $Af$ must be exactly unity, and deviation in this value (in both channels) causes $A_F$ to change.

The topology in Fig. 12-40 can be extended to more than two loops as shown in the three-channel circuit in Fig. 12-41. In this circuit, two catastrophic failures

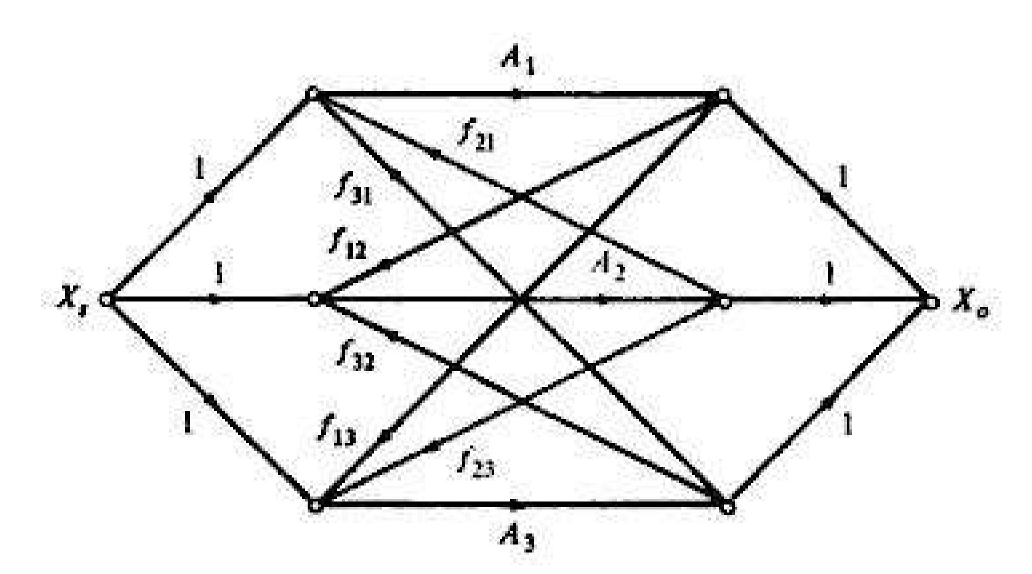



**FIGURE 12-41**
A three-channel McMillan feedback-feedforward amplifier.

and move toward each other along the negative real axis as $T_O$ is increased from zero. The poles coincide at $-(\omega_1 + \omega_2)/2$; thus, $A_F$ has two equal poles at $-(\omega_1 + \omega_2)/2$. The poles become complex as $T_O$ is increased further, with the real part remaining at $-(\omega_1 + \omega_2)/2$. We note that the closed-loop poles always remain in the left half plane and, as described in Sec. 13-2, indicate a stable system.

**The Three-Pole Function** When $T(s)$ and $A_{OL}(s)$ are three-pole functions, as is generally true for a three-stage amplifier, the problem of determining the bandwidth is still more complex. The closed-loop gain is expressible as

$$A_F(s) = \frac{A_{FO}}{1 + \dfrac{s}{1 + T_O}\left(\dfrac{1}{\omega_1} + \dfrac{1}{\omega_2} + \dfrac{1}{\omega_3}\right) + \dfrac{s^2}{1 + T_O}\left(\dfrac{1}{\omega_1\omega_2} + \dfrac{1}{\omega_1\omega_3} + \dfrac{1}{\omega_2\omega_3}\right) + \dfrac{s^3}{(1 + T_O)\omega_1\omega_2\omega_3}}$$

or

$$A_F(s) = \frac{A_{FO}}{1 + \dfrac{a_1 s}{1 + T_O} + \dfrac{a_2 s^2}{1 + T_O} + \dfrac{a_3 s^3}{1 + T_O}} \qquad (13\text{-}9)$$

The angular frequencies of the open-loop poles are $\omega_1$, $\omega_2$, and $\omega_3$ and are all on the negative real axis as displayed in Fig. 13-2. As is evident in the root locus in Fig. 13-2, increasing $T_O$ can drive two of the poles into the right half plane. These right-half-plane poles introduce terms having positive exponents in the transient response thus making the amplifier unstable. However, observe in Fig. 13-2 that when the poles remain in the left half plane so that the amplifier is stable, the decrease in the $a_1$ coefficient by $1 + T_O$ illustrates that the bandwidth is improved. In subsequent sections, we describe in detail the degree of bandwidth improvement and the costs which must be borne to obtain it.

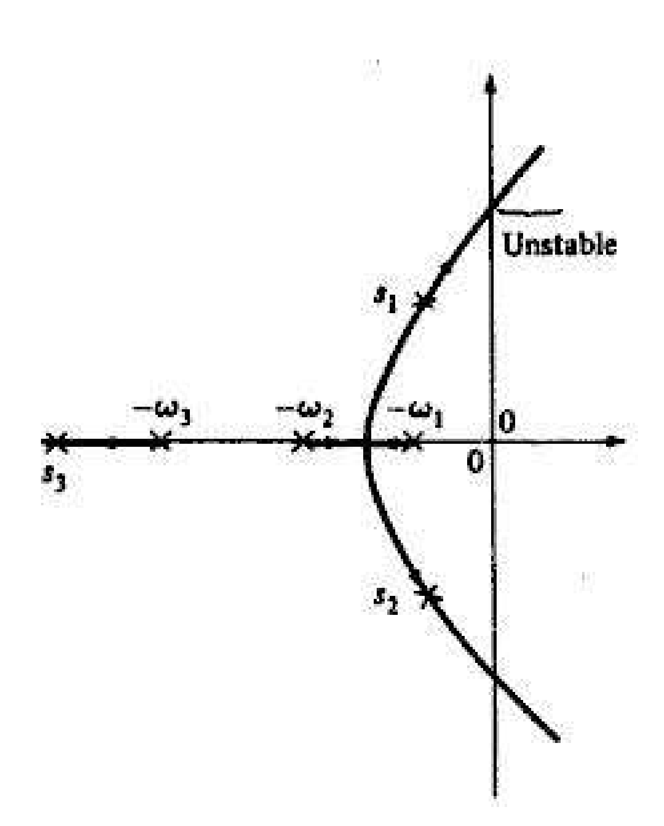



**FIGURE 13-2**
Root locus of a three-pole transfer function. The poles without feedback ($T = 0$) are $-\omega_1$, $-\omega_2$, and $-\omega_3$, whereas the poles after feedback is added ($T > 0$) are $s_1$, $s_2$, and $s_3$

and, for a stable open-loop system, the closed-loop gain $A_{F1}(s)$ has no right-half-plane poles and thus is stable.

This is not true for the Nyquist diagram in Fig. 13-3*b*, which, following the same procedure as for Fig. 13-3*a*, indicates two encirclements of $-1 + j0$. Thus $1 + T_2(s)$ has two right-half-plane zeros and $A_{F2}(s)$ is unstable as it contains two right-half-plane poles.

## Phase Margin

The Nyquist diagrams in Fig. 13-3 are redrawn for $\omega \geq 0$ in Fig. 13-4, on which the unit circle, corresponding to $T(j\omega) = 1$ (0 dB) is also constructed. The angular frequency at which the Nyquist diagram and unit circle intersect is called the *gain-crossover angular frequency* $\omega_G$ because $T(j\omega) > 1$ for $\omega < \omega_G$ and $T(j\omega) < 1$ for $\omega > \omega_G$. Comparing the two Nyquist diagrams, we observe that at $\omega_{G1}$, $\angle T_1 > -180°$, that is, $|\angle T_1| < 180°$ and the system is stable, whereas $\angle T_2(j\omega_{G2}) < -180°$ ($|\angle T_2| > 180°$) corresponds to an unstable system. It is, therefore, convenient to introduce the *phase margin* $\phi_M$, defined as

$$\phi_M \equiv \angle T(j\omega_G) + 180° \qquad (13\text{-}11)$$

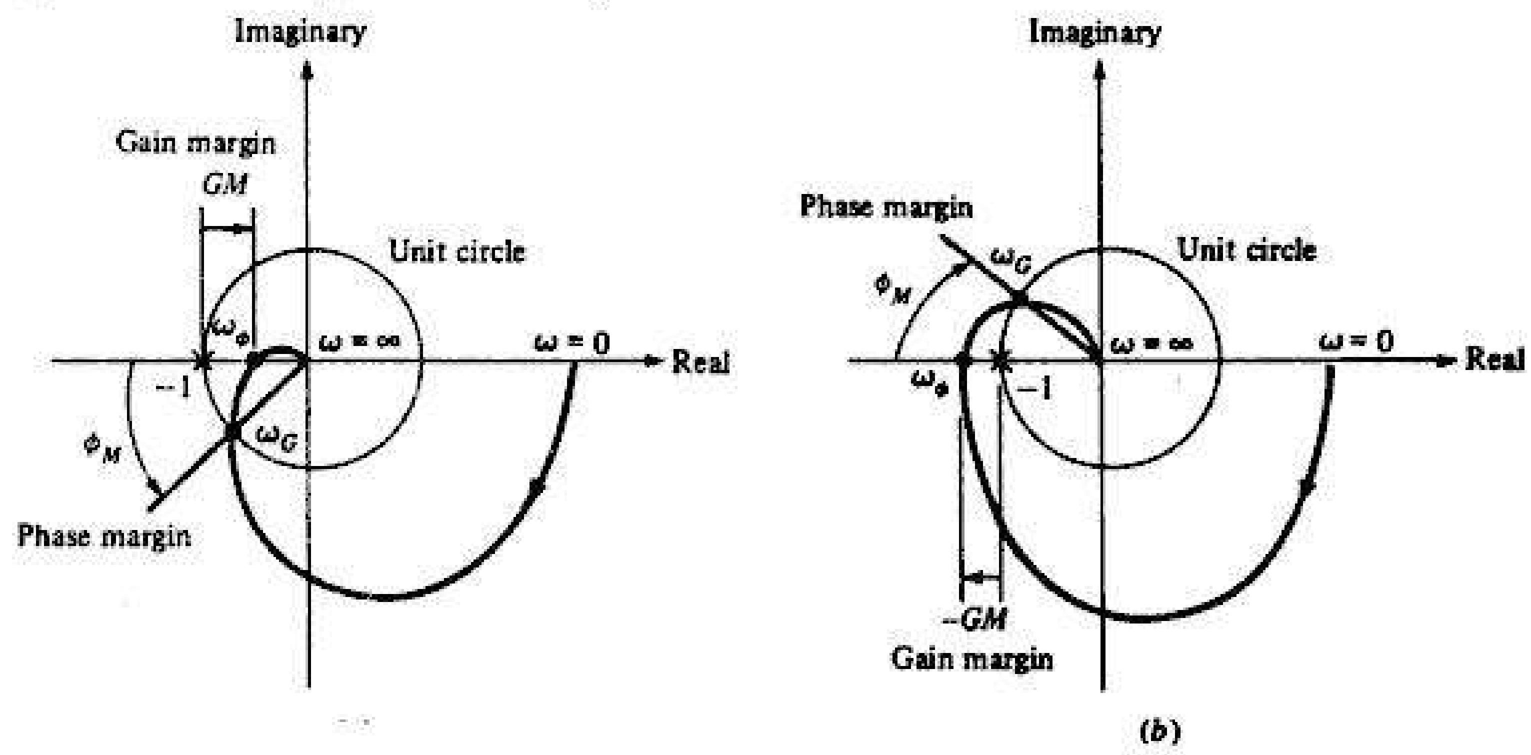
Note that $\angle T(j\omega_G)$ is, in general, a negative number. An alternative to the statement of Nyquist's criterion, extremely useful in design, is: *The closed-loop system is stable when the phase margin is positive* ($\phi_M > 0$). Hence $\angle T(j\omega_G)$ must be less negative than $-180°$.

## Gain Margin

In Fig. 13-4 the angular frequency at which the Nyquist diagram intersects the negative real axis, corresponding to $\angle T = -180°$, is defined as the *phase-crossover angular frequency* $\omega_\phi$. For $\omega > \omega_\phi$, $\angle T < -180°$ and $\angle T > -180°$ for $\omega < \omega_\phi$. The magnitude $T(j\omega_\phi)$ is used to define the *gain margin GM* as

$$GM = -20 \log T(j\omega_\phi) = -T(j\omega_\phi) \quad \text{dB} \qquad (13\text{-}12)$$

**FIGURE 13-4**
The portion of the Nyquist diagram in Fig. 13-3 for $\omega \geq 0$ used to define phase margin $\phi_M$ and gain margin $GM$. In (*a*) $\phi_M > 0$ and $GM > 0$ and the system is stable, whereas for the unstable system in (*b*) both $\phi_M$ and $GM$ are negative.



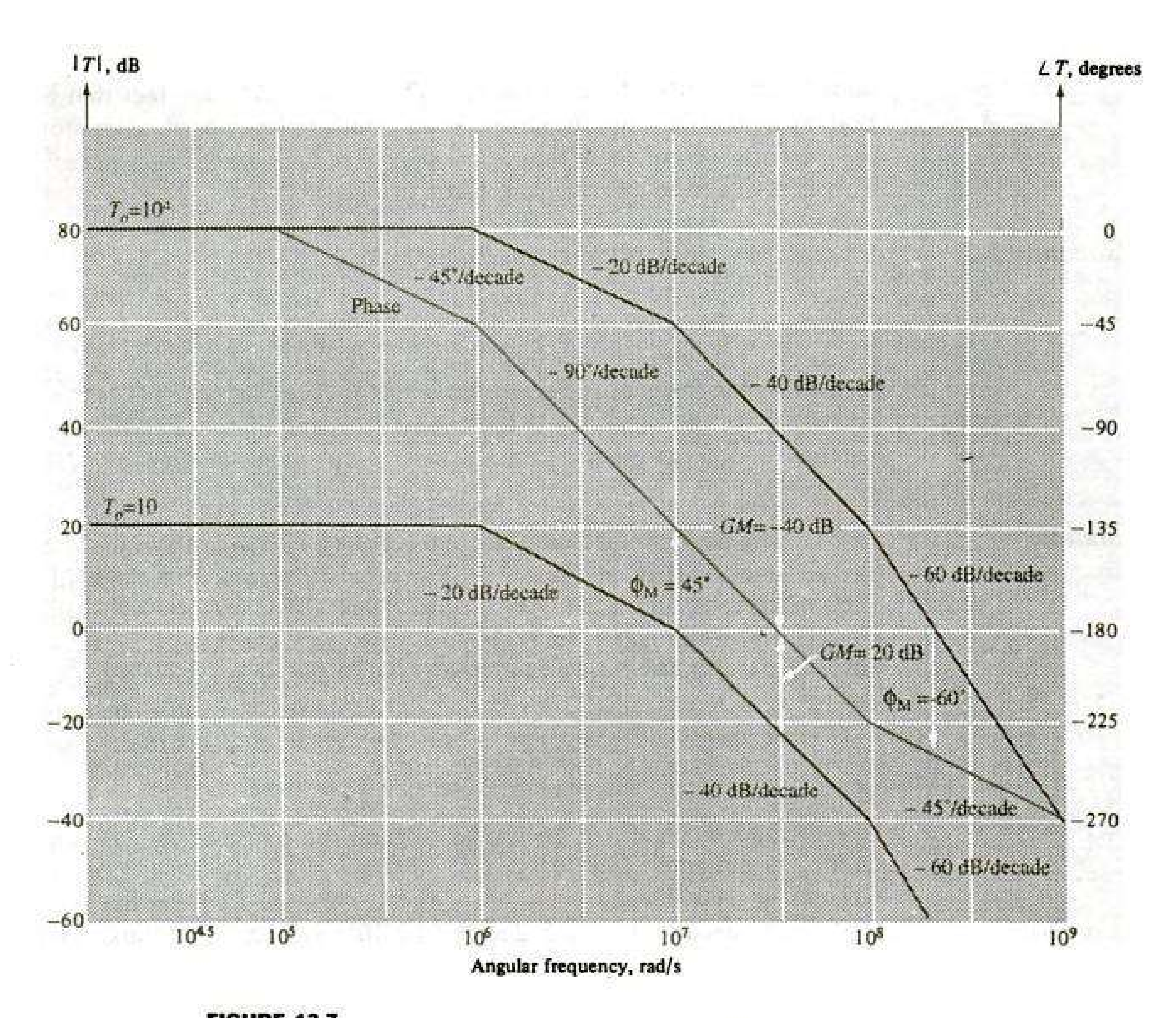



**FIGURE 13-7**
The asymptotic Bode diagram for the three-pole amplifier in Example 13-2. Note that the phase characteristic is the same for both values of $T_O$.

Several conclusions can be reached by comparing the results obtained in Example 13-2. The unstable amplifier, having $T_O = 10^4$ and $\omega_1 = 10^6$ rad/s, was stabilized by reducing either $T_O$ or $\omega_1$. The decrease in $T_O$ was accompanied by a corresponding decrease in $\omega_G$ without altering $\omega_\phi$ and thus made the amplifier stable. This, however, is an ineffective method for stabilizing the amplifier because decrease in $T_O$ also results in decreased benefits (sensitivity, distortion, etc.) derived from negative feedback.

Decrease in $\omega_1$ also results in decreased value of $\omega_G$ without change in $T_O$. For this situation $\omega_1 \ll \omega_2$ and no change in $\omega_\phi$ occurred. This is because the dominant pole ($\omega_1$) of $T(s)$ can introduce a maximum of only 90° of phase lag. Hence $\omega_\phi$ must then be due to the 90° of phase lag provided by the nondominant

where it is assumed that $(1 + T_O)/Q^2 \gg 1$. This assumption is quite reasonable as $Q^2 < 1$ for most practical responses, and $1 + T_O$ is large (at least 10) to obtain the required desensitivity, and so on. Consequently, we observe that the open-loop poles must be widely separated. Hence, although the two-pole amplifier is inherently stable, it, too, must often be compensated to achieve the desired closed-loop response.

The phase margin $\phi_M$ is obtained from $T(s)$, which can be expressed as

$$T(s) = \frac{T_O}{(1 + s/\omega_1)(1 + s/n\omega_1)} \tag{13-37}$$

The angular gain-crossover frequency $\omega_G$, obtained from Eq. (13-37) by forming $T(j\omega_G) = 1$, is

$$\frac{\omega_G}{\omega_1} = \sqrt{\frac{n^2 + 1}{2}}\left[\sqrt{\frac{4n^2(T_O^2 - 1)}{(n^2 + 1)^2} + 1} - 1\right]^{1/2} \tag{13-38}$$

For $n^2 \gg 1$ and $T_O^2 \gg 1$, which is the usual case, and use of Eq. (13-36), we can write Eq. (13-38) as

$$\frac{\omega_G}{\omega_1} = \frac{T_O}{Q^2\sqrt{2}}\left(\sqrt{4Q^4 + 1} - 1\right)^{1/2} = \frac{n}{\sqrt{2}}\left(\sqrt{4Q^4 + 1} - 1\right)^{1/2} \tag{13-39}$$

In Eq. (13-39) we note that $\omega_G$ is also widely separated from $\omega_1$ for normally encountered values of $T_O$ and $Q$.

The phase margin $\phi_M$ [given in Eq. (13-11)] for $T(s)$ in Eq. (13-37) is

$$\phi_M = -\tan\frac{\omega_G}{\omega_1} - \tan^{-1}\frac{\omega_G}{n\omega_1} + 180°$$

and can be rewritten as

$$\phi_M = \left(90 - \tan^{-1}\frac{\omega_G}{\omega_1}\right) + \left(90 - \tan^{-1}\frac{\omega_G}{n\omega_1}\right) \tag{13-40}$$
$$= \tan^{-1}\frac{\omega_1}{\omega_G} + \tan^{-1}\frac{n\omega_1}{\omega_G}$$

Since $\omega_1 \ll \omega_G$, $\tan^{-1}(\omega_1/\omega_G)$ is a very small angle and often can be neglected. Then

$$\phi_M \approx \tan^{-1}\frac{n\omega_1}{\omega_G} = \tan^{-1}\frac{\omega_2}{\omega_G} \tag{13-41}$$

and substitution of Eq. (13-39) yields

$$\phi_M \approx \tan^{-1}\sqrt{2}\,(\sqrt{4Q^4 + 1} - 1)^{-1/2} \tag{13-42}$$

Observe that Eq. (13-41) expresses the same relation as is discussed in Sec. 13-4, namely, that $\phi_M$ is determined by $\omega_2$ when the open-loop poles are widely separated.

Equation (13-15) is used to obtain the closed-loop bandwidth $\omega_H$ by solving for $\omega_H$ when $A_F(j\omega_H) = A_{FO}/\sqrt{2}$ and is

As this is more than three octaves, dominant-pole conditions apply. For these conditions, we showed in Eq. (13-3) that the closed-loop dominant pole was

$$\omega_H = |s_1| = (1 + T_O)\omega_h = (1 + T_O)\omega_1$$

or

$$|s_1| = (1 + 202) \times 0.112 \times 10^6 = 2.274 \times 10^7 \text{ rad/s}$$

Clearly, the two values are approximately equal.

Also in Sec. 13-1, it was observed that the nondominant pole was essentially unaffected by feedback when dominant-pole conditions are satisfied. Comparison of $|s_2|$ and $\omega_2$ supports this conclusion.

(*c*) From the results in part *a*, we can write

$$T(s) = \frac{T_O}{[1 + (s/\omega_1)][1 + (s/\omega_2)]}$$
$$= \frac{202}{\{1 + [s/(0.112 \times 10^6)]\}\{1 + [s/(2.354 \times 10^8)]\}}$$

The asymptotic Bode diagram for $T(j\omega)$ is displayed in Fig. 13-18*a*, from which we observe that $\phi_M = 90°$.

The asymptotic Bode diagram for $\mathbf{T}(j\omega)$ in Fig. 13-18*b* includes the effect of the third pole of the system. Using the extension of the method in Sec. 11-9, it can be shown that

$$a_3 = 1704 + 417.5C_C = 1704 + 417.5 \times 55 = 24{,}670 \text{ (ns)}^3$$

and

$$\omega_3 = \frac{a_2}{a_3} = \frac{37{,}930}{24{,}670} = 1.538 \times 10^9 \text{ rad/s}$$

Note that the inclusion of the third pole does not change $\phi_M$. (*d*) The phase margin is determined by the use of Eq. (13-42):

$$\phi_M = \tan^{-1} \sqrt{2}\,[\sqrt{4(0.3106)^4 + 1} - 1]^{-1/2} = 84.5°$$

(*e*) The results in parts *c* and *d* compare favorably. The difference is attributed to the small error introduced by the asymptotic approximation of the phase characteristic. Note that at $\omega_G = 2.263 \times 10^7$ the phase contribution of $\omega_2$ is

$$\theta_2 = -\tan^{-1}\frac{\omega_G}{\omega_2} = \frac{2.354 \times 10^6}{2.263 \times 10^7} = -5.5°$$

and this is exactly the difference between the results in parts *c* and *d*.

Computer analysis of the shunt triple in this example yields the following results: $T_O = 202$, $\omega_1 = 0.1120 \times 10^6$ rad/s, $\omega_2 = 2.406 \times 10^8$ rad/s, and closed-loop poles $s_1 = -2.87 \times 10^7$ rad/s, $s_2 = -1.76 \times 10^8$ rad/s. We observe that the open-loop poles are virtually identical to those obtained by the ap-

Using the capacitance values given in Fig. 13-19*b* and the computed values of resistance, we obtain

$$a_1 = 0.422 \times 11 + 23.4 \times 1.5 + 1.96 \times 22 + 202 \times 1.5 = 385.9 \text{ ns}$$

$$a_2 = 0.422 \times 11(2.14 \times 1.5 + 1.96 \times 22 + 202 \times 1.5) +$$
$$23.4 \times 1.5(0.179 \times 22 + 20.3 \times 1.5) + 1.96 \times 22 \times 2.02 \times 1.5$$
$$= 2959 \text{ (ns)}^2$$

Then the open-loop poles have angular frequencies of $\omega_1 = 1/a_1 = 2.59 \times 10^6$ rad/s, $\omega_2 = a_1/a_2 = 1.30 \times 10^8$ rad/s. In Example 12-8, the value of $T = 24.9$ was obtained. Use of Eqs. (13-47) and (13-35) results in

$$Q = \frac{\sqrt{2959(1 + 24.9)}}{385.9} = 0.717$$

$$s = \frac{-2.59 \times 10^6(50.3 + 1)}{2}(1 \pm \sqrt{1 - 4Q^2})$$
$$= -6.64 \times 10^7 (1 \pm j1.03) \text{ rad/s}$$

The value of $n = \omega_2/\omega_1 = 1.30 \times 10^8/2.59 \times 10^6 = 50.3$ is used in the preceding equations.

Note that the value of $Q = 0.717$ is nearly 0.707 for which no peaking in the amplitude response exists. Thus we can conclude that the amplitude response of this series-shunt pair exhibits virtually no peaking. Similarly, the step response has little overshoot (<3 percent).

(*b*) The asymptotic Bode diagram for

$$T(s) = \frac{24.9}{(1 + s/2.59 \times 10^6)(1 + s/1.30 \times 10^8)}$$

is displayed in Fig. 13-21. The phase margin, indicated in Fig. 13-21, is $\phi_M = 60.5°$. If Eq. (13-42) is used, $\phi_M = 65.0°$. The difference between the two results is attributed to the errors in the asymptotic approximation of the phase characteristic in Fig. 13-21. The error in phase due to the pole at $-\omega_2$ is approximately 5° (about the difference between the two values).

Computer analysis of the circuit in Fig. 13-19 gives

$$\omega_1 = 2.64 \times 10^6 \text{ rad/s} \qquad \omega_2 = 1.27 \times 10^8 \text{ rad/s}$$

$$s = -6.59 \times 10^7 (1 \pm j\,1.035) \text{ rad/s}$$

The approximate values are all within 3 percent of actual (computer-simulated) values and demonstrate the effectiveness of the approximate calculations as a design tool.

Now let us examine some of the reasons underlying the values listed in Table 14-1. Since the basic Op-Amp stages are feedback amplifiers, high open-loop gain—and hence large return ratios—is desirable to ensure exclusive dependence of the closed-loop gain on the feedback resistances $R_1$ and $R_2$ (see Fig. 10-42). Similarly, most Op-Amps are designed to have a transfer function that contains a dominant pole. Under these conditions, both the open- and closed-loop gain-bandwidth products are equal. Thus, for a specified closed-loop gain, the closed-loop bandwidth is readily determined.

A high common-mode rejection ratio *CMRR* is needed to ensure that the output signal is proportional to the difference between the input voltages. With high *CMRR* values, common-mode signals, often containing dc components, have only a small effect on the amplifier output.

To approximate the characteristics of an ideal voltage amplifier, the Op-Amp must possess high-input and low-output resistances. The current in each of the two inputs is ideally zero. These dc currents are part of the bias currents of the input stage and, because ICs are invariably direct-coupled, must be small to prevent adverse interaction with the signal source.

In an "ideal" IC, we can exactly match transistor characteristics and component values. Regardless of how sophisticated the fabrication technology is, exact matching is impossible in the "real world." The offset voltage and current are measures of the degree of mismatch in the circuit, and clearly these should be small.

**The Two-Stage Architecture** A majority of commercially available operational amplifiers employ the structure displayed in Fig. 14-1. This cascade configuration is commonly referred to as a *two-stage operational amplifier* because only the differential amplifier and gain stage contribute to the overall voltage gain. The differential amplifier is used as the input stage to provide the inverting and noninverting inputs, the high *CMRR*, and the high-input resistance, as well as voltage gain. The low-output resistance of the Op-Amp is achieved by the emitter-follower output stage. The level shifter adjusts the dc voltages so that the output voltage signal is referenced to ground. The adjustment of dc levels is required because the gain stages are direct-coupled. Since large-valued capacitors cannot be fabricated on a chip, virtually all ICs are direct-coupled. The interior gain stage is a high-gain voltage amplifier used to obtain a large open-loop gain.

In the previous paragraph we note that the input and output stages are required to match the Op-Amp with the "external world." That is, these stages serve as the interfaces between the amplifier and the input signal sources and

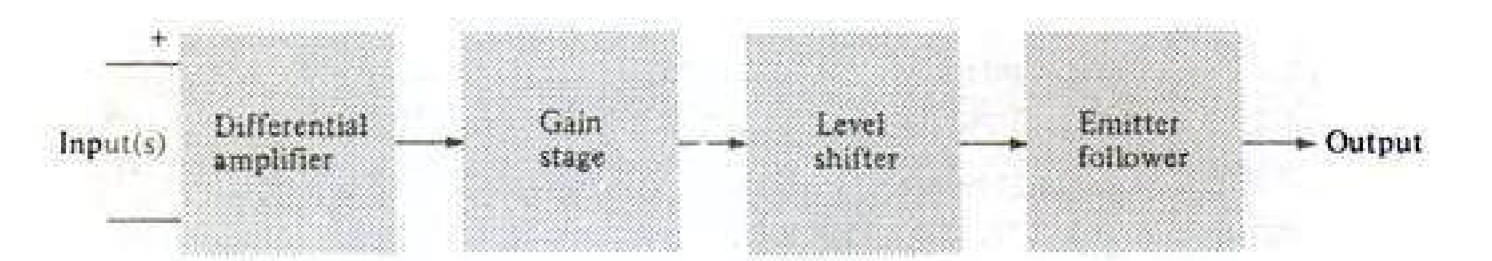



**FIGURE 14-1**
The architecture of a two-stage Op-Amp.

However, limitations such as frequency response often preclude Darlington pair input stages. The use of super-$\beta$ transistors (Sec. 5-3) in the differential pair results in high input resistances at current levels usually encountered. For a super-$\beta$ transistor ($\beta_o$ = 5000) biased at $I_C$ = 12.5 $\mu$A, $R_{id}$ = 20 M$\Omega$. Clearly, reduction in the current results in increased $R_{id}$.

**The Differential-Mode Gain $A_{DM}$** Since the input stage of the Op-Amp is one of the two gain stages, it is desirable to make $A_{DM}$, the differential-mode gain, large. Consequently, active loads are also employed in these stages. From Eq. (10-86), we obtain

$$|A_{DM}| = \left|\frac{v_o}{v_{DM}}\right| = \frac{\beta_o R_L}{r_\pi} = g_m R_L \tag{14-6}$$

Recalling that $v_{DM} = V_d/2$ [Eq. (10-91)], we can express the differential-output voltage of the amplifier as

$$V_o = \frac{g_m R_L}{2} V_d \tag{14-7}$$

where $R_L$ is the parallel combination of the active-load resistance $R_C$ and the output resistance $r_o$ of the active element. Expression of $g_m$ in terms of the bias current yields

$$|V_o| = \frac{I_C}{2V_T}|V_d|\, R_L = \frac{I_O}{4V_T}|V_d|\, R_L \tag{14-8}$$

Equation (14-8) indicates that the effective transconductance of a differential stage is one-fourth that of a single BJT biased by a collector current of $I_O$.

*Example 14-2* The input stage of the 741 Op-Amp is depicted in Fig. 14-9. The *npn-pnp* transistor combinations *Q*1 and *Q*3 and *Q*2 and *Q*4 form the active element of the differential pair. The active loads are provided by the three-transistor current source *Q*5, *Q*6, and *Q*7. Transistors *Q*8 and *Q*9 form a *pnp* current mirror used for base-biasing and ensure that the transistors remain in the active region when no input signal is applied. Transistors *Q*1 to *Q*6 are biased at $I_C$ = 9.5 $\mu$A, and all have $\beta_o$ = 250. The Early voltages are 100 and 50 V for the *npn* and *pnp* transistors, respectively.

Determine the gain $V_o/V_d$, the differential-input resistance $R_{id}$, and the output resistance $R_o$. Use the small-signal schematic representation of the differential stage in Fig. 14-10.

*Solution* A convenient method for analyzing this circuit is to obtain the Norton equivalent of the stage. Conversion of the Norton equivalent to a Thévenin equivalent gives both the output resistance and the voltage gain. In Fig. 14-10, KCL requires that $I_o = I_{c4} + I_{c6}$. The composite active element *Q*2 and *Q*4 can be considered as an emitter follower (*Q*2) driving *Q*4, connected as a common-base stage, that is, a cascode circuit. This is illustrated in the equivalent circuit

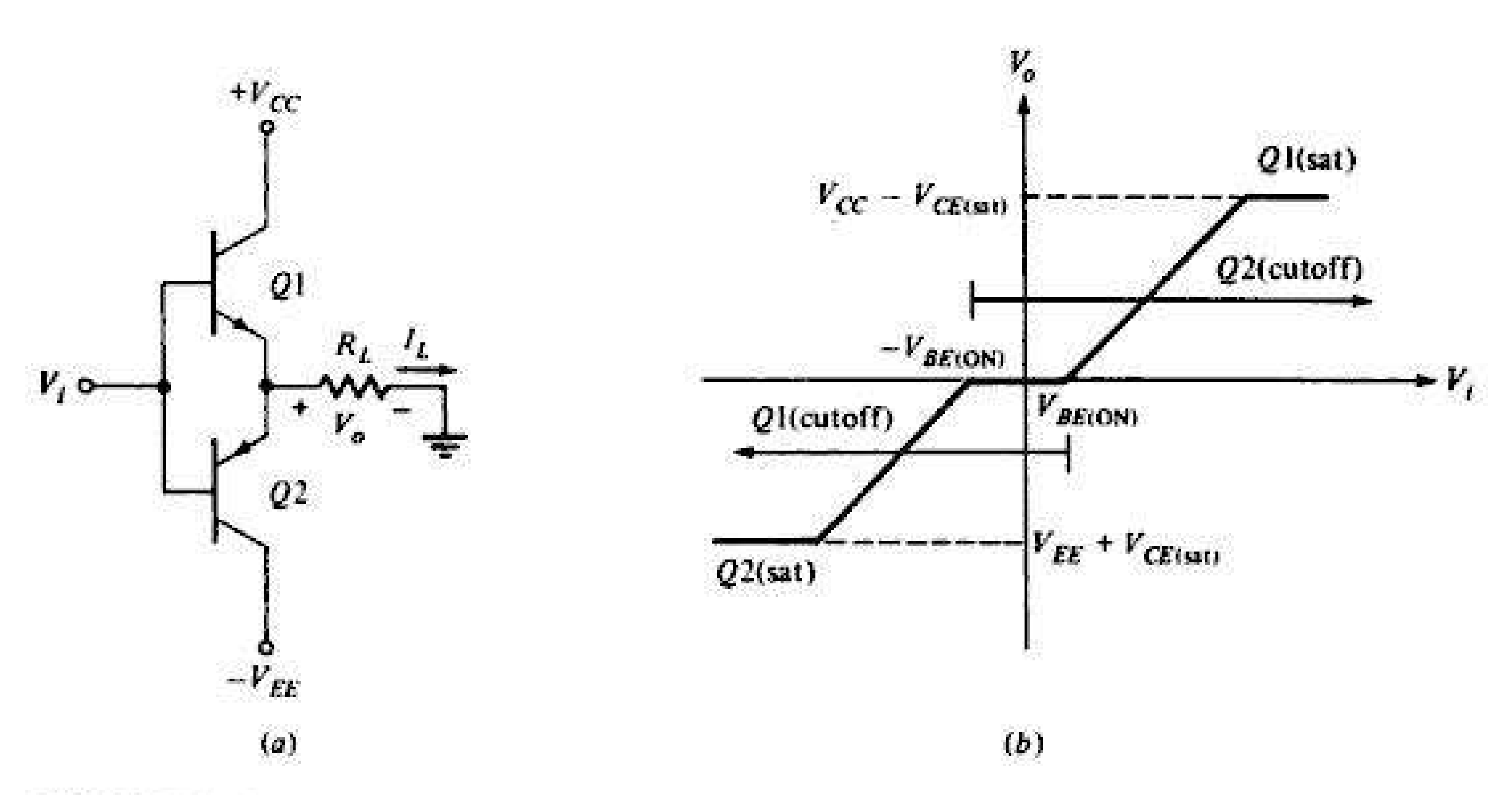



**FIGURE 14-15**
Complementary emitter-follower output stage: (*a*) Circuit diagram, (*b*) voltage transfer characteristic. The horizontal portion of the characteristic in the vicinity of the origin introduces crossover distortion.

There is a fundamental difficulty with the circuit in Fig. 14-15*a* because the output voltage remains virtually zero until $V_i = V_{BE(ON)}$. This phenomenon is known as *crossover distortion* and is readily observed in the transfer characteristic illustrated in Fig. 14-15*b*. [Actually, the output becomes nonzero for $V_i \approx V_{BE(\text{cut in})} = V_\gamma \approx 0.5$ V. However, the current in the transistor is so small that the output voltage is negligible.]

Crossover distortion can be virtually eliminated by applying a bias voltage $V > 2V_\gamma$ between the two bases so that a small current exists in the transistors in the quiescent state. One common technique is to employ a pair of series-connected *pn* diodes as displayed in Fig. 14-16. It is common to fabricate diodes $D1$ and $D2$ as diode-connected BJTs (Sec. 5-6). The transfer characteristic of the circuit in Fig. 14-16*a* is displayed in Fig. 14-16*b*, in which we see that the crossover distortion is virtually eliminated. However, the characteristic does not pass through the origin, and, with $V_i = 0$, $V_o \neq 0$. Recalling that $V_i$ is obtained from the level-shift stage, we obtain $V_o = 0$ with zero input signal by making the quiescent value of $V_i \approx -V_{BE2}$.

The circuit in Fig. 14-17 is also used to eliminate crossover distortion. The block labeled $V$ is the $V_{BE}$-multiplier circuit in Fig. 14-14*b* and is used in place of diodes $D1$ and $D2$ in Fig. 14-16*a*. The output voltage of this block is designed to apply approximately 1.1 V between the two bases. Hence both $Q1$ and $Q2$ are conducting slightly under quiescent conditions.

The output stages in Figs. 14-16*a* and 14-17 are both employed in commercially fabricated 741-type Op-Amps. The basic output-stage configuration is displayed in Fig. 14-18. Transistors $Q14$ and $Q20$ form the complementary emitter follower. The small resistances $R_6$ and $R_7$ provide for current-limiting

since $R' \gg R$. The difference voltage $V_d$ is the voltage $-V_i$ across $R_i$. If we take the input offset voltage into account (Fig. 14-20b) and use Eqs. (14-17) and (14-12) with $R' \gg R$, we obtain

$$V_d = 2V_{DM} = V_1 - V_2 - V_{io} = -\frac{VR}{R'} - V_{io} = -\frac{R}{R'}(V + V_3) \qquad (14\text{-}19)$$

and

$$V_{CM} = \tfrac{1}{2}(V_1 + V_2) = V_s + \frac{VR}{2R'} \qquad (14\text{-}20)$$

Substitution of Eqs. (14-19) and (14-20) into Eq. (14-17) yields

$$-A_{DM}\frac{R}{2R'}(V + V_3) + A_{CM}\left(V_s + \frac{VR}{2R'}\right) = 0 \qquad (14\text{-}21)$$

Since $A_{DM} \gg A_{CM}$ the fourth term in this equation may be neglected compared with the first term. Hence, if the measured value of $V$ is designated by $V_6$, we obtain

$$CMRR\,\frac{R}{2R'}(V_6 + V_3) = V_s \qquad (14\text{-}22)$$

For $CMRR = 10^5$, $R/R' = 2 \times 10^{-3}$, and $V_s = 10$ V we find that $V_6 + V_3 = 0.1$ V. For $V_{io} = 5$ mV, $V_1 = 5$ V. Hence $V_6 = -4.9$ V, and very poor accuracy is obtained from this measurement since two large and almost equal voltages $|V_6|$ and $|V_3|$ must be subtracted. This difficulty is overcome by changing the input to a new value $V_s'$ and measuring the new value of $V$, called $V_6'$. Then, corresponding to Eq. (14-22), we have

$$CMRR\,\frac{R}{2R'}(V_6' + V_3) = V_6' \qquad (14\text{-}23)$$

Subtraction of Eq. (14-22) from Eq. (14-23) results in elimination of $V_3$ and yields

$$CMRR = \frac{2R'}{R}\,\frac{V_s' - V_s}{V_6' - V_6} \qquad (14\text{-}24)$$

If $V_s' = 5$ V, $V_s = -5$ V, $CMRR = 10^5$, and $R'/R = 500$, we obtain $V_6' - V_6 = 0.1$ V. However, this subtraction can now be done electronically with the SHS subtract circuit in Fig. 14-24. The switch $S_3$ is closed for the $V_6$ measurement and opened for the $V_6'$ measurement.

## 14.8 FREQUENCY RESPONSE AND COMPENSATION

The closed-loop response of the basic inverting and noninverting Op-Amp stages often is required to exhibit dominant-pole behavior for all values of the low-frequency closed-loop gain. Thus the Op-Amp can be represented by a one-pole transfer function. In previous sections we demonstrated that the open-loop poles must be widely separated and the phase margin $\phi_M \approx 90°$ to achieve this kind of closed-loop

The maximum current that can be delivered by the differential stage is twice the quiescent collector current observed in the transfer characteristic of the emitter-coupled pair (Fig. 3-38). Thus

$$\frac{dV_o}{dt} = \text{slew rate} = \frac{2I_C}{C_C} \tag{14-31}$$

For a type 741 Op-Amp ($I_C \approx 9.5\ \mu$A and $C_C = 30$ pF), the slew rate is $2 \times 9.5/30 = 0.63$ V/$\mu$s.

Substitution of Eq. (14-27) into Eq. (14-31) results in

$$\text{Slew rate} = \frac{8\pi I_C}{g_m} f_G \tag{14-32}$$

and, as $g_m = I_C/V_T$, Eq. (14-32) becomes

$$\text{Slew rate} = 8\pi V_T f_G \tag{14-33}$$

In Eq. (14-33) we observe that the slew rate is increased by increasing $f_G$, the unity gain bandwidth of the Op-Amp. However, increases in $f_G$ are limited by the frequency response of the transistors used. For a two-stage Op-Amp, such as a type 741, $f_G$ can be increased only marginally since $f_T$ of the lateral *pnp* transistors is of the order of 5 to 10 MHz. Significant improvement in $f_G$ is usually achieved by employing three-stage architectures. The National Semiconductor LM118 is one such amplifier and has $f_G = 15$ MHz and a slew rate of 50 V/$\mu$s.

Improvement in the slew rate can also be achieved by decreasing $g_m$ for a given $f_G$. We have seen previously that the use of an emitter resistance (emitter degeneration) decreases the gain of a common-emitter stage (Sec. 10-11). Consequently, the effective value of $g_m$ for the stage is reduced for a given bias current and results in an increased slew rate. Employment of emitter degeneration in the input stage of a type 741 Op-Amp improves the slew rate by almost an order of magnitude.

**Effect of Slew Rate on an Input Signal** Let us consider a unity gain noninverting Op-Amp stage biased by $\pm 15$-V supplies to which a 15-V step input voltage is applied. The Op-Amp has a slew rate of 0.5 V/$\mu$s. Since the output voltage $V_o$ cannot change by more than 0.5 V/$\mu$s, the output waveform is as shown in Fig. 14-34. As indicated in Fig. 14-34, $V_o$ does not reach 15 V, the expected output, until 30 $\mu$s have elapsed.

Now, let us consider that the input voltage $V_s$ to this Op-Amp stage is $V_s = V_m \sin \omega t$. With no slew-rate limitation, $V_o = V_m \sin \omega t$ and

$$\frac{dV_o}{dt} = \omega V_m \cos \omega t \tag{14-34}$$

The imbalance varies the input signals to $A_1$, thus changing the reference current to the current sources $I_1$ and $I_2$. The output connections of $A_1$ are such that when an imbalance exists, the output tends to adjust itself and make $I_{C1} = I_{C2}$. The difference between $I_1$ and $I_2$, from KCL, is the current $I_G$ in the gain resistance $R_G$. Similarly, the current sources $I_3$ and $I_4$ are imbalanced and a difference current $I_s$ exists in the sense resistance $R_S$. The action of $A_2$ when the circuit becomes imbalanced is analogous to that of $A_1$; thus, the output of $A_2$ tends to render $I_{C3} = I_C$. Since $V_o$ is the difference between the base voltages of $Q3$ and $Q4$

$$V_o = I_S R_S \tag{14-41}$$

The currents $I_S$ and $I_G$ are equal, a consequence of using matched differential current sources. Combination of Eqs. (14-40) and (14-41) gives

$$\frac{V_o}{V_{\text{in}}} = A_V = \frac{R_S}{R_G} \tag{14-42}$$

Clearly, adjustment of the resistance ratio permits the realization of different values of $A_V$.

**The Operational Transconductance Amplifier** A voltage-to-current converter (Sec. 12-1) is an amplifier that produces an output current that is proportional to an input voltage. The constant of proportionality is referred to as the *transconductance* of the amplifier. An OTA is a single-chip amplifier in which the transconductance is controlled by an externally connected resistance.

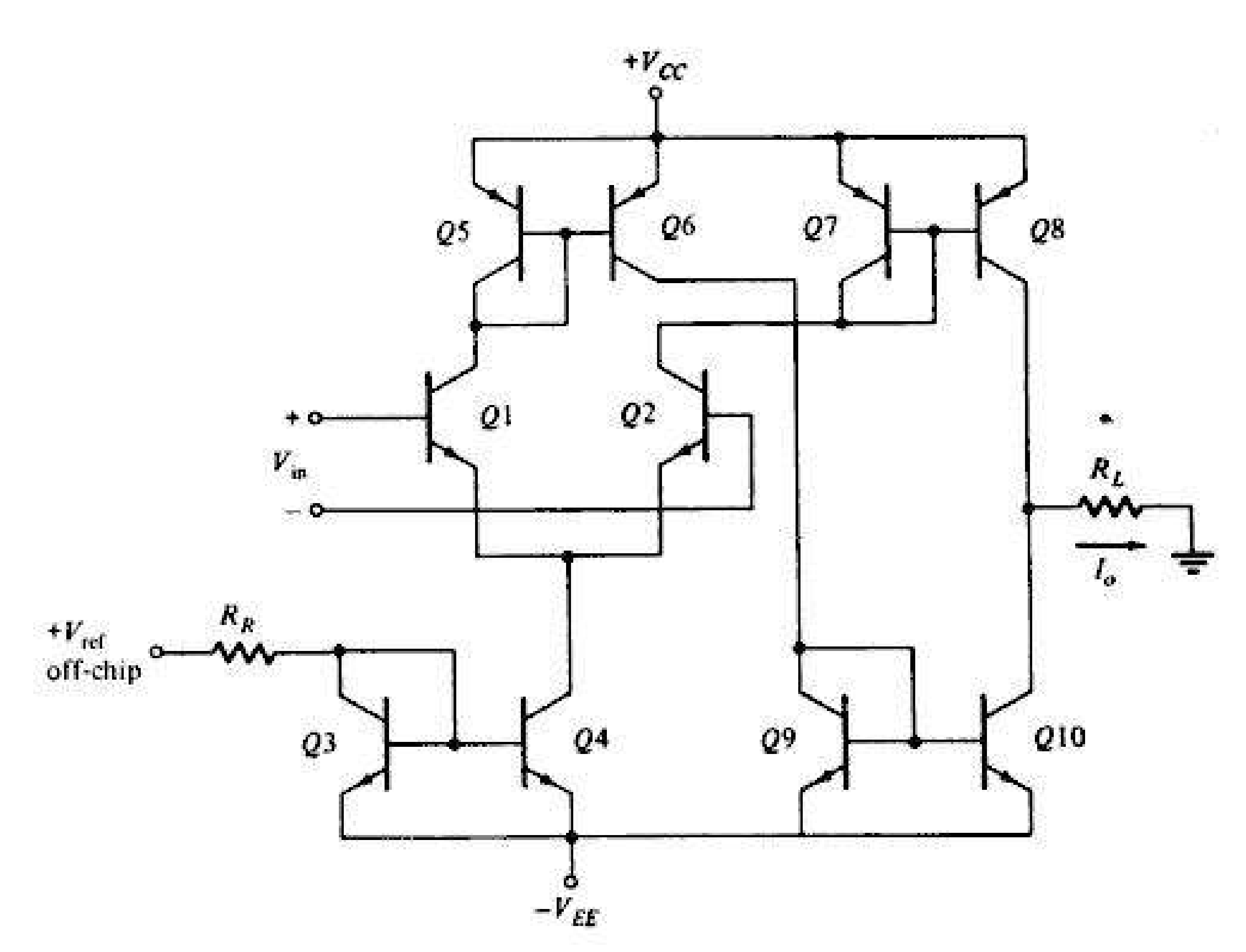


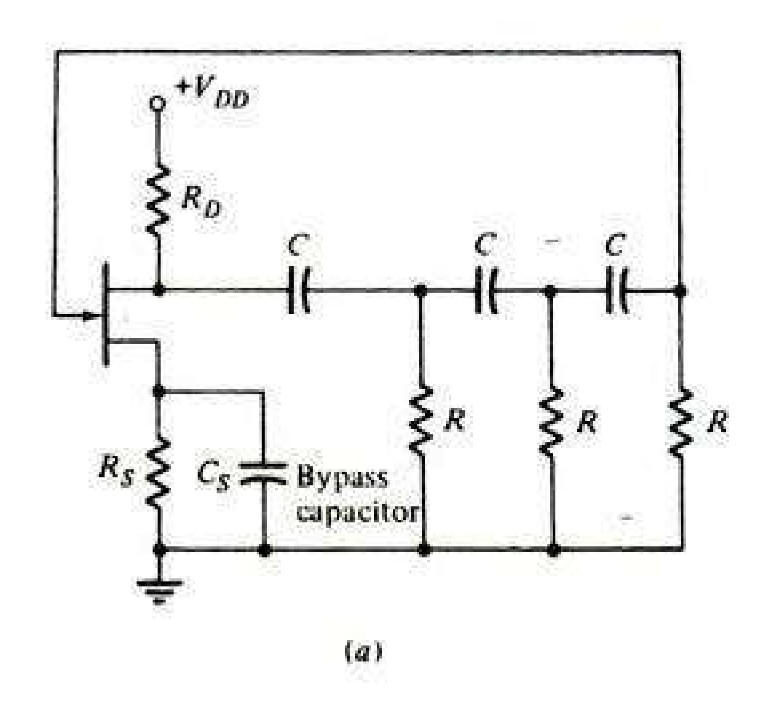

**FIGURE 14-42**
An operational transconductance amplifier.

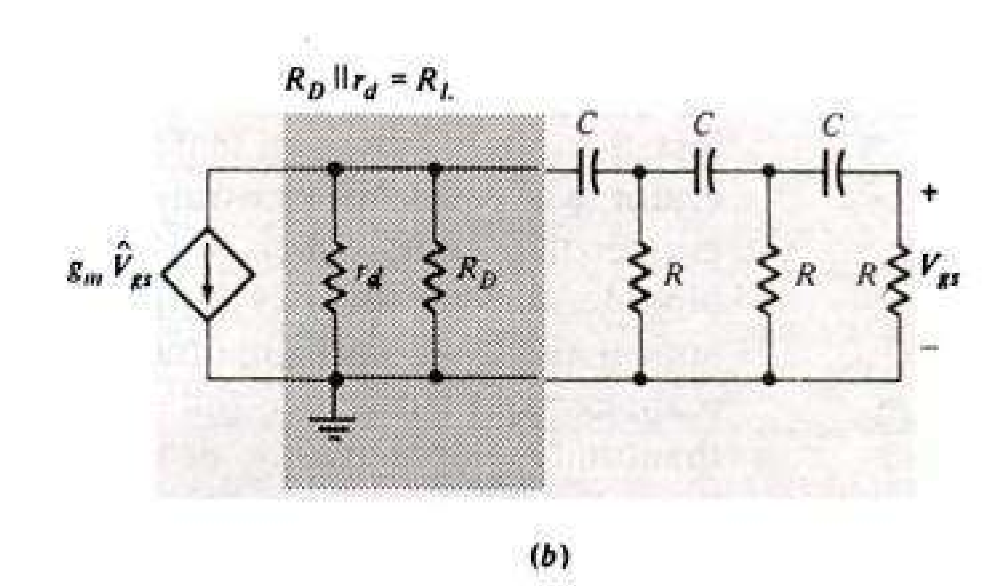
*Part Four*

# SIGNAL PROCESSING AND DATA ACQUISITION

The transmission, reception, and processing of information in the form of electrical signals is the basis of modern electronic systems for control, communication, and computation. Many of these systems utilize both analog and digital signals to perform their function. Clearly, a variety of different signal waveforms are required. Furthermore, the form of these signals (amplitude, phase, frequency, duration, rise time, etc.) must be appropriate for the particular application for effective processing. In the two chapters in this part of the book we treat a number of different circuits used for signal generation and processing. Chapter 15 deals with waveform generation and waveshaping. Included are sinusoidal oscillators, clock (square-wave) generators, and time-base generators. The conversion of data from analog-to-digital (A/D) and digital-to-analog (D/A) form is described in the second chapter of this part. In addition, signal-conditioning circuits such as logarithmic amplifiers, integrators, multipliers, and active filters are discussed. The circuits treated in this section utilize the basic building blocks (logic gates, Op-Amps, switches, etc.) described in the previous parts of the book.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.



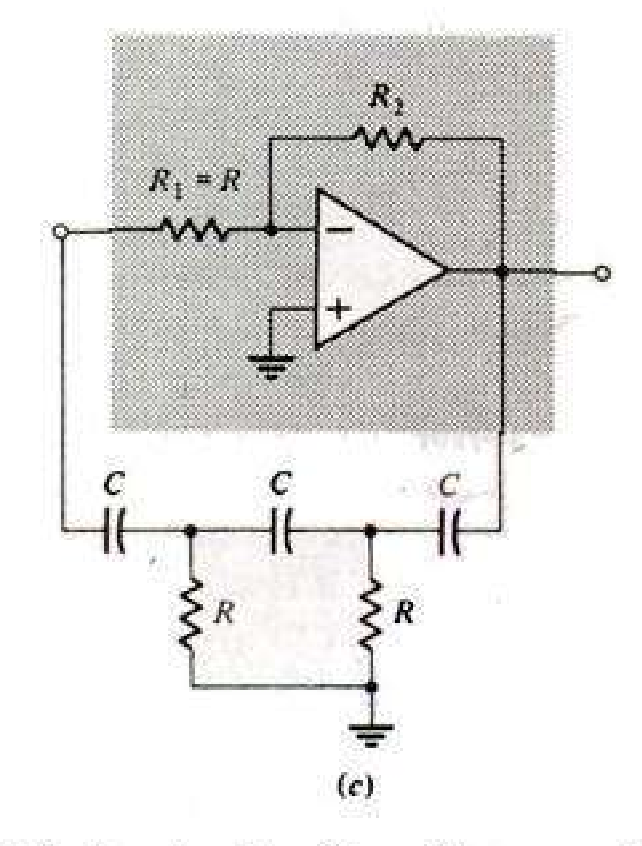




**FIGURE 15-2** (*a*) A JFET *RC* phase-shift oscillator and (*b*) its equivalent circuit. (*c*) An Op-Amp version of the phase-shift oscillator.

which the circuit will oscillate, provided that the magnitude of the amplification is sufficiently large.

Determination of $\mathbf{T}(j\omega)$ by the method described in Sec. 12-7 (Prob. 15-1) gives

$$\mathbf{T}(j\omega_N) = \frac{jg_m R_L \omega_N^3}{(1 - 6\omega_N^2) + j\omega_N(5 - \omega_N^2)} \tag{15-4}$$

where $\omega_N = \omega RC$. Application of Eq. (15-3) yields $\omega_N^2 = \frac{1}{6}$, from which the frequency of oscillation $f_o$ becomes

$$f_o = \frac{1}{2\pi RC\sqrt{6}} \tag{15-5}$$

At $\omega_N = 1/\sqrt{6}$,

$$\mathbf{T}(j\omega_N) = \frac{jg_m R_L/6\sqrt{6}}{j(5 - \frac{1}{6})/\sqrt{6}}$$

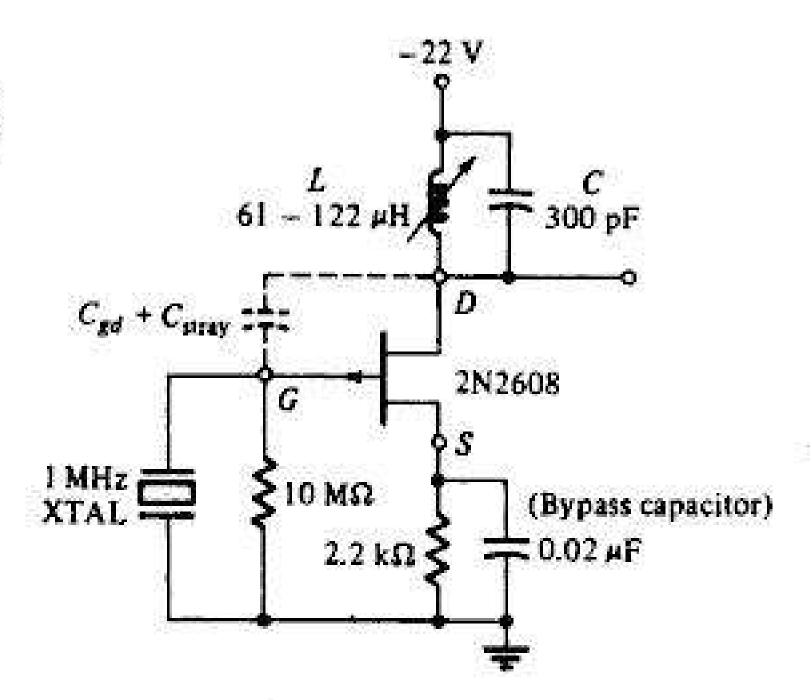



**FIGURE 15-8**
A 1-MHz FET crystal oscillator. (*Courtesy of Siliconix Co.*)

reactance is inductive, and outside this range it is capacitive, as indicated in Fig. 15-7.

A variety of crystal-oscillator circuits is possible. If a crystal is used for $Z_1$ in the basic configuration in Fig. 15-5*a*, a tuned *LC* combination for $Z_2$, and the capacitance $C_{gd}$ between gate and drain for $Z_3$, the resulting circuit is as indicated in Fig. 15-8. From the theory given in the preceding section, the crystal reactance, as well as that of the *LC* network, must be inductive. For the loop gain to be greater than unity, we see from Eq. (15-13) that $X_1$ cannot be too small. Hence the circuit will oscillate at a frequency which lies between $\omega_s$ and $\omega_p$ but close to the parallel-resonance value. Since $\omega_p \approx \omega_s$, the oscillator frequency is essentially determined by the crystal, and not by the rest of the circuit.

## 15-6 MULTIVIBRATORS

The oscillators described in the previous sections are one part of the class of *regenerative circuits*. We observed that at the frequency of oscillation, sinusoidal oscillators are positive-feedback amplifiers. Multivibrators are another important group of regenerative circuits that are used extensively in timing applications. Multivibrators are conveniently classified as (1) bistable circuits, (2) monostable circuits, or (3) astable circuits.

The bistable latches and flip-flops described in Secs. 7-1 to 7-3 can all be considered as *bistable multivibrators*. In Sec. 15-11 we describe the regenerative comparator (Schmitt trigger), another bistable circuit. One important characteristic of a bistable circuit is that it maintains a given output state (level) unless an external signal (trigger) is applied. Application of an appropriate external signal causes a change of state, and this output level is maintained indefinitely until a second trigger is applied. Thus a bistable circuit requires two external triggers before it returns to its initial state.

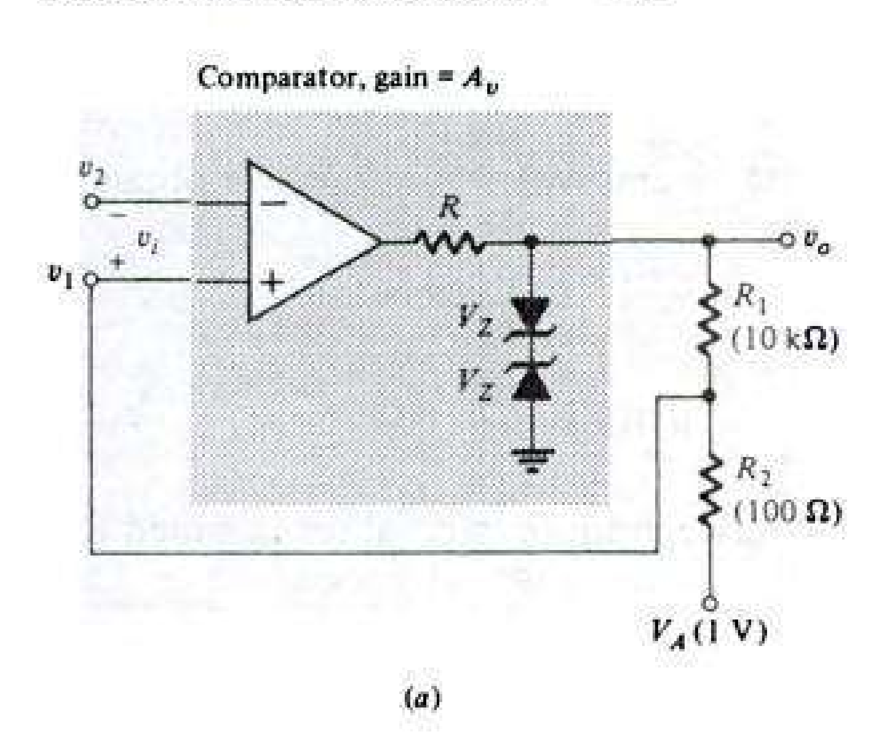
The *monostable*, or "one-shot," multivibrator generates a single pulse of specified duration in response to each external trigger signal. As its name implies, only one stable state exists. Application of a trigger causes a change

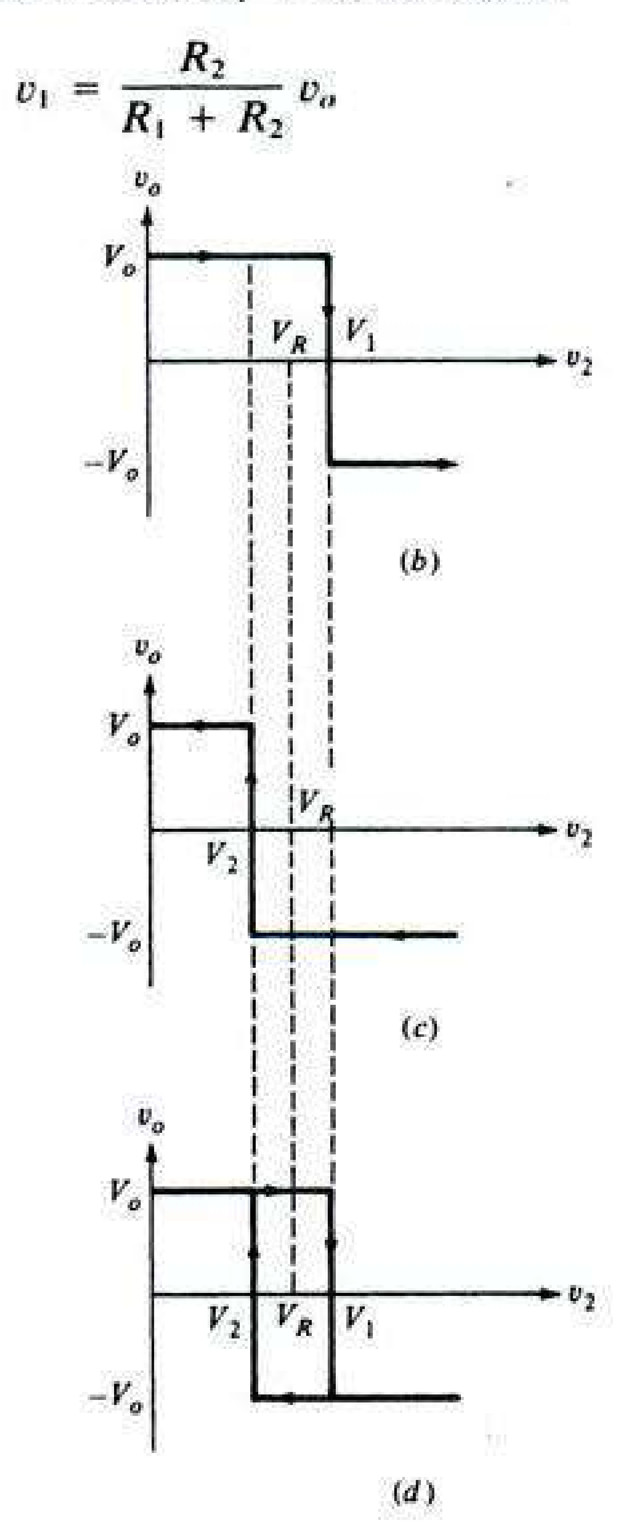
**15-9 REGENERATIVE COMPARATOR (SCHMITT TRIGGER)** The transfer characteristic in Fig. 15-20*b* makes the change in output from −7 to +7 V for a swing in input of about 1.0 mV. Hence, the voltage gain is 14,000. By employing positive feedback the gain may be increased greatly. Consequently the total output excursion takes place in a time interval during which the input is changing by much less than 1 mV. Theoretically, if the return ratio $T$ is adjusted to be −1, then the gain with feedback $A_V$ becomes infinite [Eq. (12-4)]. Such an idealized situation results in an abrupt (zero rise time) transition between the extreme values of output voltage. If a loop gain in excess of unity is chosen, the output waveform continues to be virtually discontinuous at the comparison voltage. However, the circuit now exhibits a phenomenon called *hysteresis*, or *backlash*, which is explained in the following.

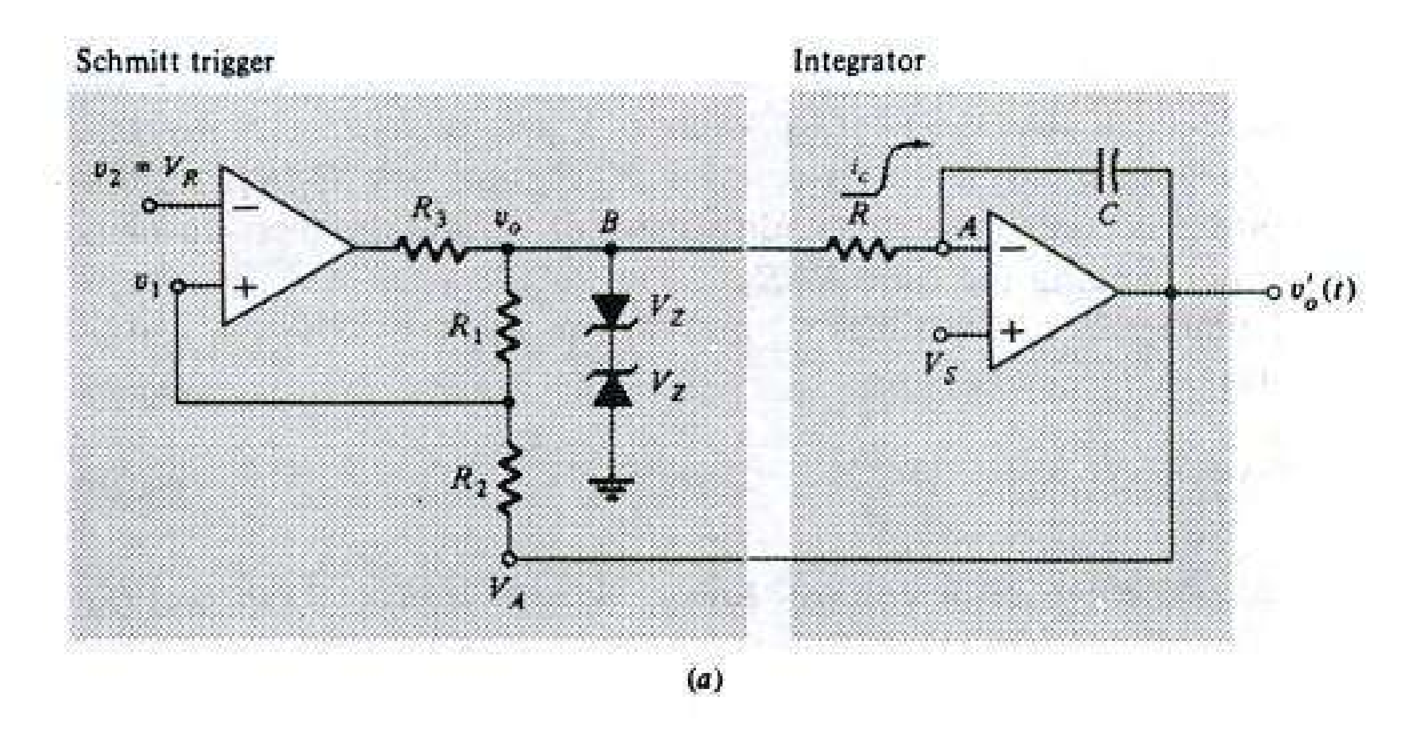
The regenerative comparator of Fig. 15-22*a* is commonly referred to as a *Schmitt trigger* (after the inventor of a vacuum-tube version of this circuit). The input voltage is applied to the inverting terminal 2 and the feedback voltage to the noninverting terminal 1. Assuming that the output resistance of the comparator is negligible compared with $R_1 + R$, we obtain

$$v_1 = \frac{R_2}{R_1 + R_2} v_o$$

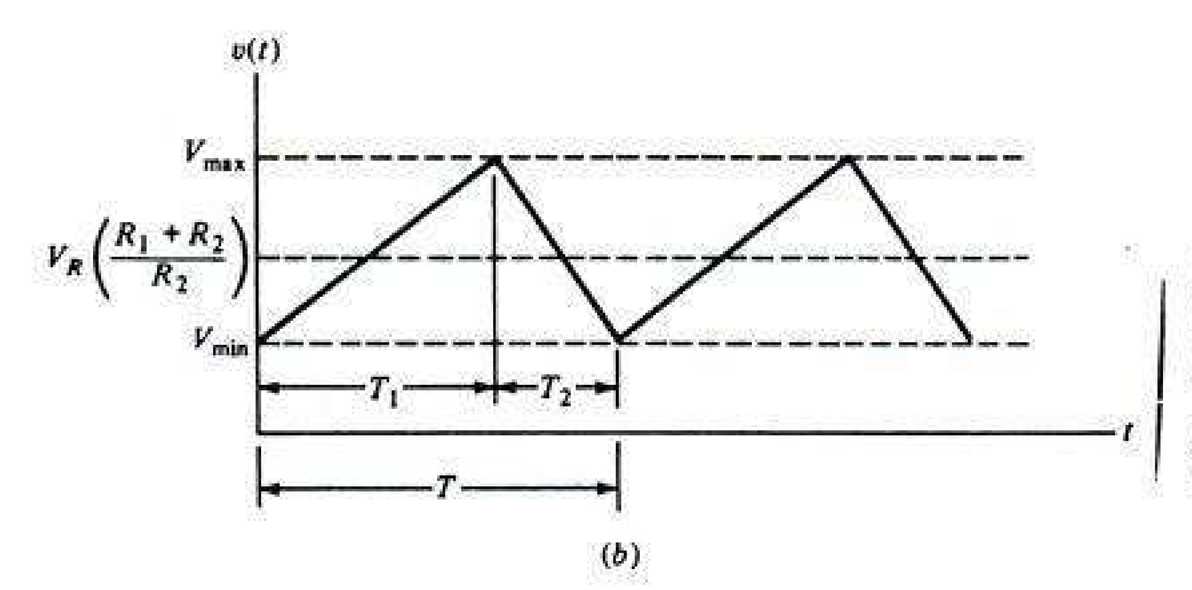
**FIGURE 15-22**
(*a*) The regenerative comparator or Schmitt trigger. The output waveforms showing a transition (*b*) from $+V_o$ to $-V_o$ and (*c*) from $-V_o$ to $+V_o$. (*d*) The output voltage for one cycle showing the hysteresis ($V_1 - V_2$).






(a)




(b)

**FIGURE 15-27**
(a) A triangle-wave generator. (b) Output waveform. [*Note*: $T_1 = T_2$ if $V_S = 0$. Also, $V_{max} = V_o R_2/R_1 = -V_{o(min)}$ if $V_R = 0$. The square-wave output is $-V_o$ during the interval $T_1$ and $+V_o$ for the interval $T_2 - T_1$.]

With a negative input, the output $v'(t)$ of the integrator is an *increasing* ramp. The voltage at the noninverting comparator input $v_1$ is obtained by the use of superposition and is

$$v_1 = -\frac{V_o R_2}{R_1 + R_2} + \frac{v_o' R_1}{R_1 + R_2} \tag{15-26}$$

When $v_1$ rises to $V_R$, the comparator changes state, $v_o = +V_o$, and $v_o'(t)$ starts *decreasing* linearly. Hence the peak $V_{max}$ of the triangular waveform occurs for $v_1 = V_R$. From Eq. (15-26),

$$V_{max} = V_R \frac{R_1 + R_2}{R_1} + V_o \frac{R_2}{R_1} \tag{15-27}$$

By a similar argument it is found that

$$V_{min} = V_R \frac{R_1 + R_2}{R_1} - V_o \frac{R_2}{R_1} \tag{15-28}$$

The peak-to-peak swing is

$$V_{max} - V_{min} = 2V_o \frac{R_2}{R_1} \tag{15-29}$$

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

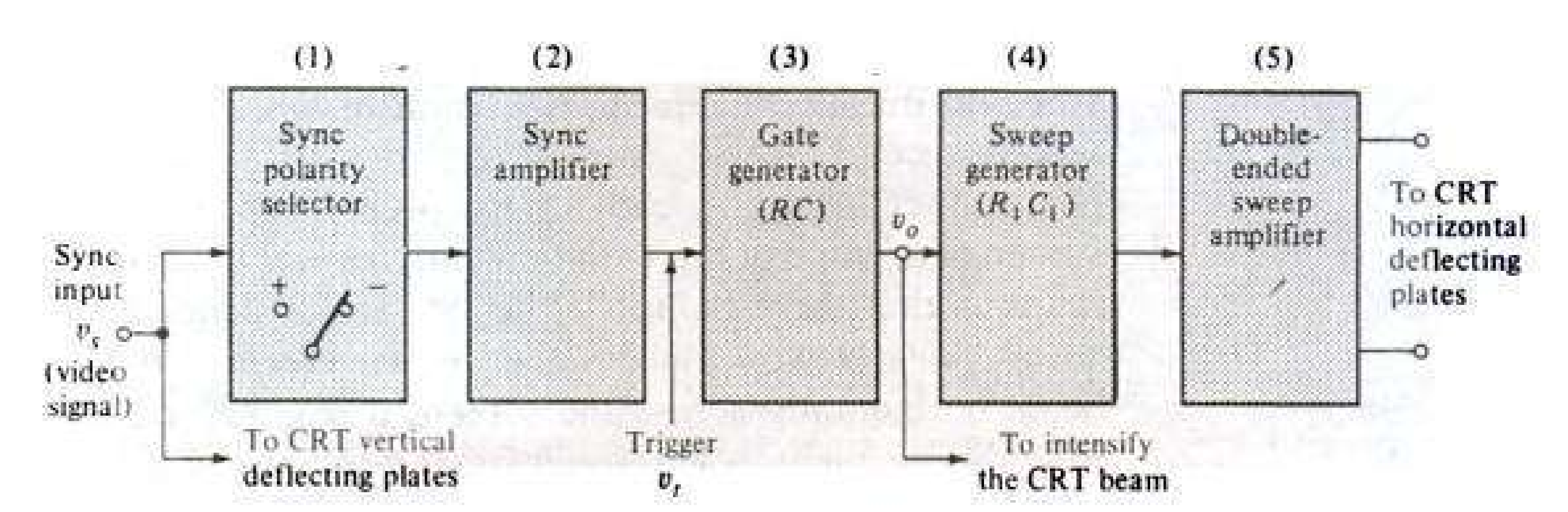



**FIGURE 15-34**
A block diagram of the time-base generating system for a cathode-ray tube (CRT).

consist of 1-$\mu$s pulses with a time interval of 100 $\mu$s between pulses. In this case the fastest recurrent sweep which will provide a synchronized pattern will have a period of 100 $\mu$s. If, typically, the time base is spread out over 10 cm, the pulse will occupy 1 mm and none of the detail of form of the pulse will be apparent. If, on the other hand, a sweep of period 1 $\mu$s or somewhat larger could be used, the pulse would be spread across the entire screen. Therefore, what is required here is a sweep set for, say, a 1.5-$\mu$s interval which remains quiescent until it is initiated by the pulse. Such a monostable circuit is known as a *driven* sweep or a *triggered* sweep.

A block diagram for a time-base system for a CRT is indicated in Fig. 15-34. The waveform $v_s$ to be observed is applied through a high-quality video amplifier (not indicated in Fig. 15-34) to the vertical deflecting plates of the CRT. This signal is simultaneously applied to the sweep system as the synchronizing input. In block 1 the sync polarity selection is made by taking the output across either a collector or an emitter resistor. The sync amplifier (block 2) need not operate linearly, since all that is required is that the output $v_t$ be large and fast enough to be able to trigger the (monostable) gate generator. In some scopes a Schmitt trigger is used to obtain a sharp pulse on either the rising or falling portion of the signal, as desired. Since this trigger is used to start the sweep, a selected portion of the input signal appears on the scope face.

The third block in Fig. 15-34 is a monostable multivibrator whose gate width is determined by the time constant $RC$ (Fig. 15-29*a*). A negative gating waveform ($v_o$ in Fig. 15-29*b*) is applied to the sweep generator (4), whose sweep speed depends upon a resistor $R_1$ and a capacitor $C_1$ (Fig. 15-35). The sweep generator output is amplified linearly (5) and applied to the horizontal deflecting plate of the CRT.

In a case in which the sweep time is short in comparison with the time between sweeps the CRT beam will remain in one place most of the time. If the intensity is reduced to prevent screen burns, the fast trace will be very faint. To intensify the trace during the sweep, a positive gate which is derived from the outputs of the multi is applied to the CRT grid. As a matter of fact, in the presence of this "unblanking" or "intensifier signal" the beam brightness may be adjusted so that the spot is initially invisible but the trace will become visible as soon as the sweep starts.

We note that when $S_1$ is open, the signal $v$ reproduces the input signal $v_m$. As we have drawn the figure, a perceptible voltage change takes place in $v_m$ during any interval when $S_1$ is open. Thus, when $v_m$ is positive, the positive extremities of the waveform $v_m$ are not at a constant voltage and, similarly, for the negative extremities when $v_m$ is negative. More customarily, the frequency of operation of the switch is very large (typically 100 times), in comparison with the frequency of the signal $v_m$. Therefore, no appreciable change takes place in $v_m$ during the interval when $S_1$ is open. Accordingly, it is proper to describe the waveform $v_m$ as a square wave of amplitude proportional to $v_m$ and having an average value (shown dashed) that is also proportional to the signal $v_m$. Alternatively stated, the waveform $v$ is a square wave at the switching frequency, amplitude-modulated by the input signal, and superimposed on a signal which is proportional to the input signal $v_m$ itself.

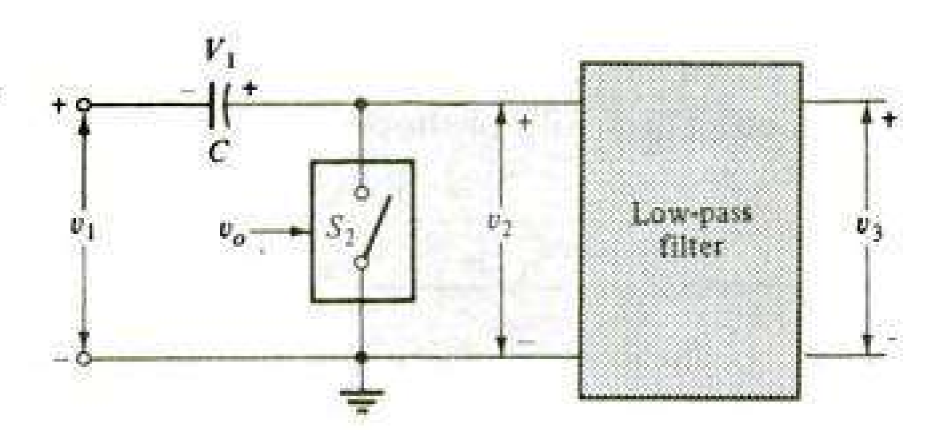
The low-frequency cutoff of the high-pass filter is such that the relatively high frequency square wave passes with small distortion while the signal frequency is well below the cutoff point. Consequently, at the output of the filter, we obtain the waveform $v_1$ in Fig. 15-41*c*, which corresponds to $v$ but with the average value subtracted. Note that $v_1$ is an attenuated replica of the amplitude-modulated waveform $v'$ obtained in Fig. 15-40.

## Demodulators

The process of recovering the modulating signal $v_m$ from the PAM waveform is called *demodulation*. The positive-negative gain amplifier, which was used as a modulator, functions equally well as a demodulator. This statement is justified by the following argument. If the modulated waveform $v'$ of Fig. 15-40*c* is used as the input $v_m$ to Fig. 15-39, then, in the interval $T_1$ (Fig. 15-40*b*), when $v' = -v_m$, the gain $A$ is $-1$, and in the next half-period $T_2$, when $v' = v_m$, $A = +1$. Hence the output $v'$ (in Fig. 15-39) in any interval is $v_m$ (in Fig. 15-40). Clearly, we have reconstituted the original signal $v_m$.

An alternative demodulator, corresponding to the chopper modulator of Fig. 15-41 is indicated in Fig. 15-42, where switch $S_2$ is controlled by $+v_o$ and, hence, operates in synchronism with switch $S_1$ in Fig. 15-41. For example, in the interval $T_1$ of Fig. 15-41*c*, $S_2$ is closed and the output is zero. Hence during $T_1$, the negative extremity of $v_1$ is clamped to ground, and the voltage across $C$ is $-V_1$, as indicated in Fig. 15-42. In the next half-cycle $T_2$ of the square wave, $S_2$ is open, $v_1 = +V_2$, and $v_2 = V_2 + V_1$, which is the amplitude of $v$ (Fig. 15-41) during $T_2$. As a consequence of the clamping action of $C$ and the controlled switch $S_2$, the waveform $v_1$ is reconverted into the chopped mod-

**FIGURE 15-42**
A synchronous demodulator.



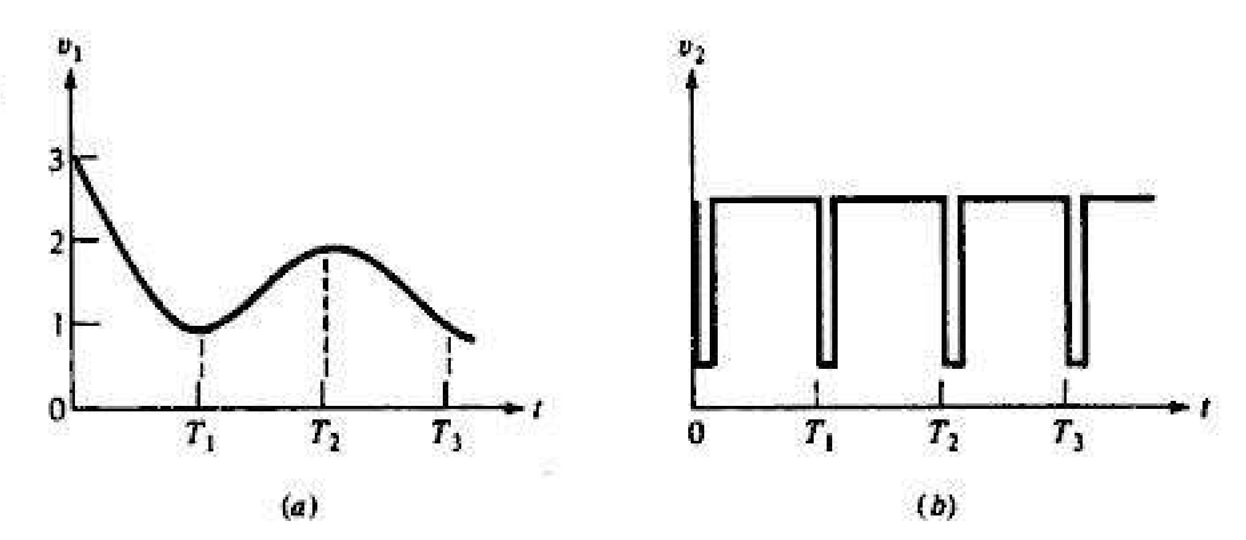


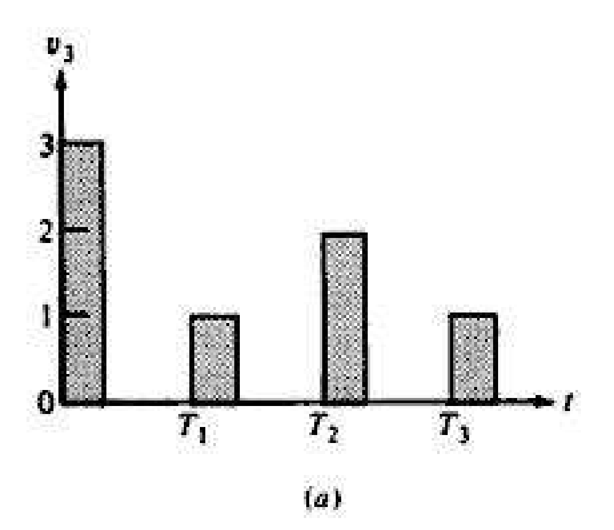

**FIGURE 16-1**
Two waveforms of continuous signals.

or simply a *sampled signal*. Sys'ems which utilize such signals care called *sampled-data systems*.

In the context of this discussion, the sequence of pulses in each time interval of Fig. 16-2*b* is a numeric, or digital, representation of the corresponding voltage samples shown in Fig. 16-2*a*. The waveforms in Fig. 16-1*b* may be the clock signal which sets the timing sequence used in the generation of the pulses in Fig. 16-2. Neither the amplitude or time of the signals in Figs. 16-1*b* and 16-2*b* correspond to the physical quantity $v_1$. Essentially, these are signals in which the information is contained by the presence or absence of a pulse during a given time interval.

The waveforms in Figs. 16-1*a* and 16-2*a* are analog signals, and those depicted in Figs. 16-1*b* and 16-2*b* are digital signals. Both types of signals are often present in modern electronic systems. Clearly, circuits which process these signals and convert one type to the other are required. The following qualitative descriptions help to indicate the several circuit functions that must be performed.

The pictorial representation of a commercial amplitude-modulation radio system is shown in Fig. 16-3. The primary purpose of the system is to transfer the audio information at the transmitting end to the receiving end. The first step in the process is to convert the acoustic energy into an electrical signal. The conversion is effected by a transducer, usually a microphone. As the output of the transducer is a low-level signal, amplification is necessary. Radio-frequency (rf) signals (signals whose frequencies are greater than 500 kHz) are



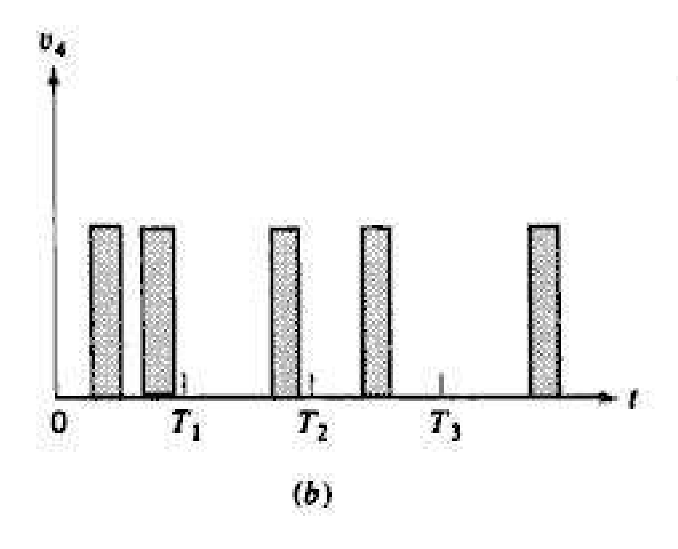




**FIGURE 16-2**
Two discrete signals. The waveform in (*a*) represents pulses whose amplitudes are those in Fig. 16-1*a* at times (0) $T_1$, $T_2$, and $T_3$, respectively. The waveform in (*b*) is the 2-bit binary representation of the amplitudes in (*a*).

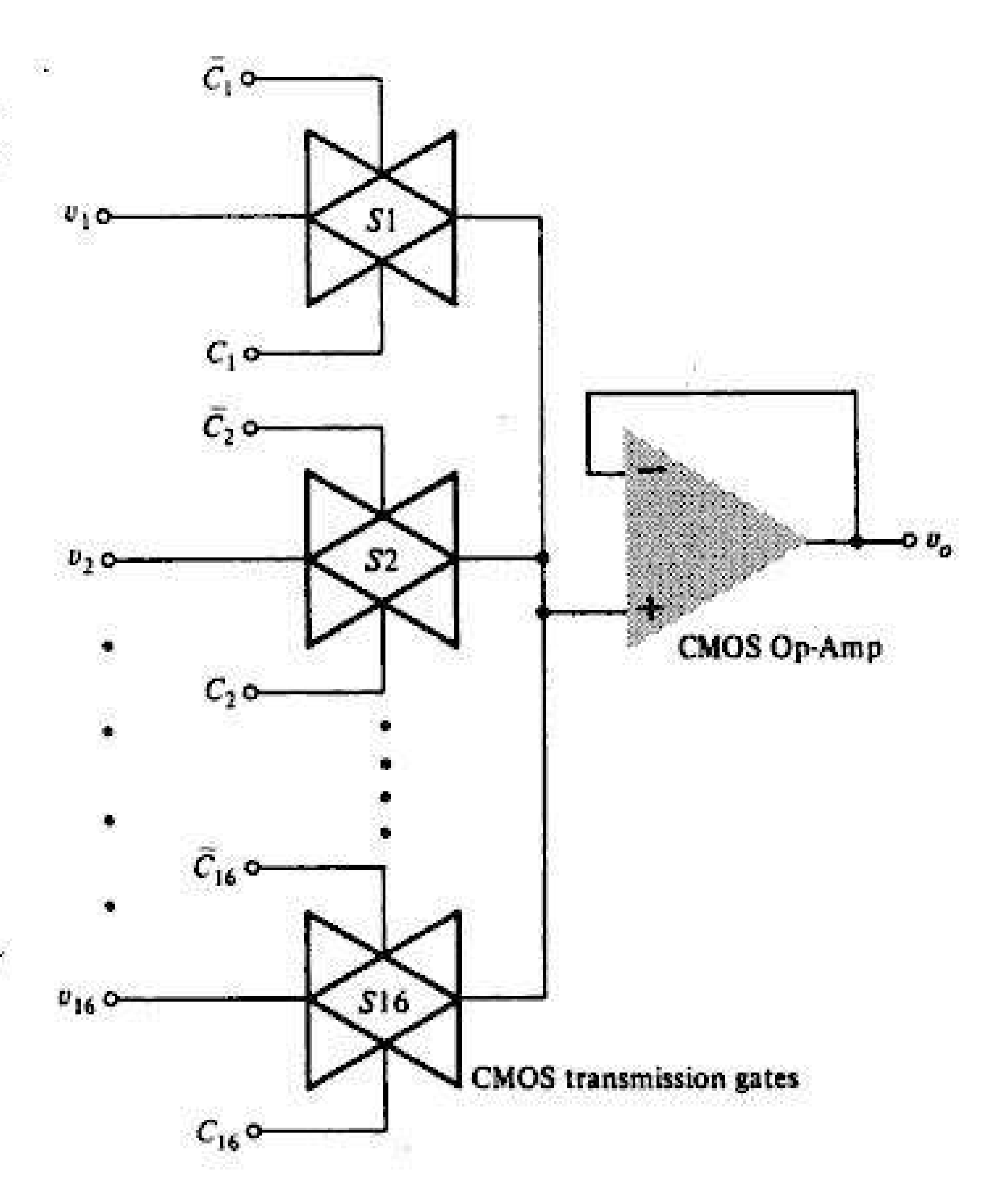



**FIGURE 16-9**
A 16-input analog multiplexer using CMOS transmission gates.

A block diagram of a 16-input analog commutator is indicated in Fig. 16-9. Time-division multiplexing results if the complementary MOSFET switch *S*1 closes (i.e., it is in its low-resistance state) for a time *T*, switch *S*2 closes for the second interval *T*, *S*3 transmits for the third period *T*, and so forth. In Fig. 16-9 the symbol $C_k$ ($k = 1, 2, \ldots, 16$) represents the digital control voltage and $\overline{C}_k$ is its complementary value obtained from an inverter (not shown). If $C_k$ equals binary 1, the CMOS gate transmits the analog signal $v_k$ to the output, but if $C_k$ is binary 0, no transmission is allowed.

The block diagram for obtaining the required digital control voltages for the analog multiplexer in Fig. 16-9 is indicated in Fig. 16-10. The control $C_k$ is the output of the *k*th line of a 4-to-16-line decoder (Sec. 7-6). The four address lines *A*, *B*, *C*, and *D* are the outputs from a binary counter which is excited by a pulse generator. If the time interval between pulses is *T*, time-division multiplexing is obtained with the system shown in Figs. 16-10 and 16-9 (cor-

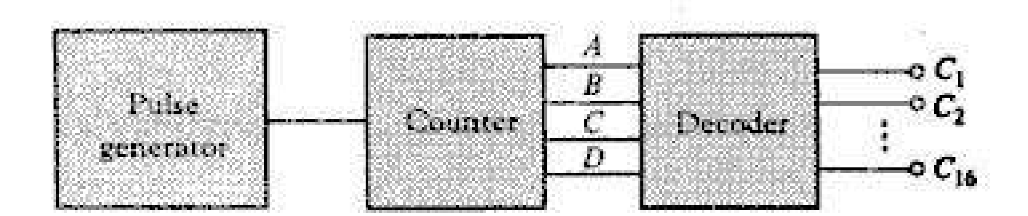



**FIGURE 16-10**
A system for generating the digital control voltages $C_k$ for the multiplexer.

between the stopband and cutoff frequencies ($f_S - f_H$ or, alternatively, $f_S - f_C$) is referred to as the *transition band*. Often $f_S$ is chosen to be one octave above the cutoff frequency ($f_S = 2f_H$). Typically, attenuation in excess of 24 dB at $f_S = 2f_H$ is commonplace.

The band-pass response in Fig. 16-27*b* shows two stopbands and two transition bands, one each above and below the cutoff frequencies $f_2$ and $f_1$ which define the passband. Similar characteristics for realistic high-pass and band-reject filters can be drawn and corresponding identifications of the various frequency bands made.

**Filter Specification** We use the low-pass response in Fig. 16-27*a* to illustrate the information that must be available to the designer of a filter. At a minimum, the designer requires the following specifications:

1. The cutoff frequency $f_H$ (or $f_C$), that is, the range of passband frequencies.
2. The stopband attenuation $H_o - H_2$.
3. The stopband frequency range. That is, $f_S$ is specified.
4. The allowable passband ripple $\gamma = H_o - H_1$. If no ripple is permitted, $\gamma = 0$ and $H_1 = H_o$.

An other factor that is usually specified is the impedance level at both the filter input and output (the interfaces with the signal source and load). The characteristic illustrated in Fig. 16-27*a* is the magnitude of the transfer function $H(j\omega)$. Often, the phase response (delay) and transient response (rise time, overshoot) of the filter are also specified.

Since commercially available Op-Amps have unity gain bandwidths in excess of 100 MHz, it is possible to design filters up to frequencies of several megahertz.[1] Because of slew rate limitations and unit-to-unit variation in Op-Amp gain-bandwidth product and open-loop gain, however, most IC active filters are used at audio frequencies. Clearly, decrease in the unity gain bandwidth of the Op-Amp results in decrease of the maximum filter frequency.

## 16-9 BUTTERWORTH AND CHEBYSHEV FILTER FUNCTIONS

The frequency responses displayed in Fig. 16-27 are approximations of the ideal low-pass and band-pass characteristics in Figs. 16-26*a* and 16-26*b*, respectively. Similar approximations for the high-pass and band-elimination characteristics in Figs. 16-26*c* and 16-26*d* can be made. To design a "real" filter, the characteristics in Fig. 16-27 must be expressed mathematically. The general form of the transfer function $H(s)$ can be expressed as

$$H(s) = \frac{A(s)}{B(s)} \tag{16-13}$$

[1]At these frequencies, Op-Amps are usually costly and alternative filter realizations are often used.

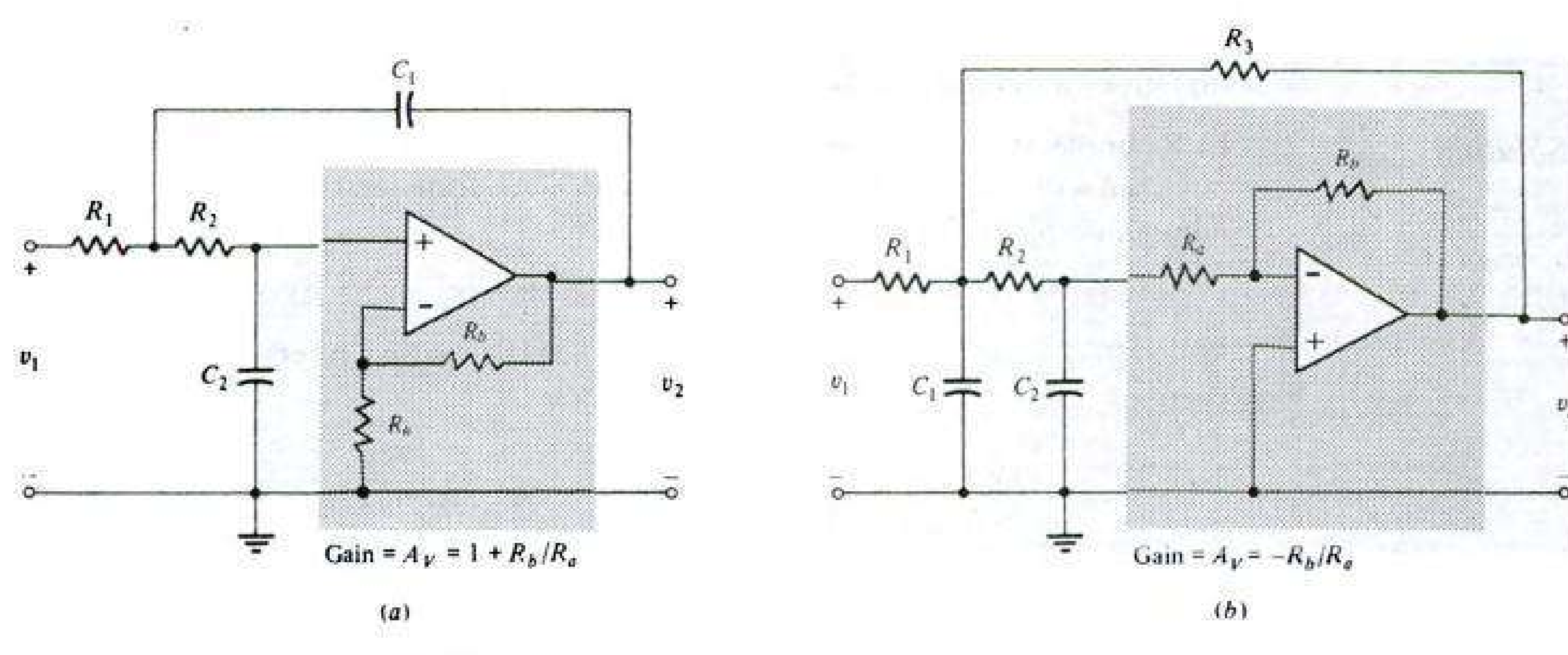



**FIGURE 16-30**
Sallen and Key low-pass sections using (*a*) a noninverting amplifier (positive-feedback) and (*b*) an inverting amplifier (negative feedback).

**Low-Pass Sections** The Sallen and Key circuit in Fig. 16-30*a* uses a noninverting Op-Amp stage to provide positive feedback. The transfer function of this circuit can be expressed as (Prob. 16-21)

$$H(s) = \frac{A_V}{R_1R_2C_1C_2s^2 + s\,[C_2(R_1 + R_2) + R_1C_1(1 - A_V)] + 1} \qquad (16\text{-}24)$$

where $A_V = 1 + R_b/R_a$ is the gain of the Op-Amp stage. Comparison of Eq. (16-24) with the low-pass function in Eq. (16-15) gives

$$\omega_o = \frac{1}{\sqrt{R_1R_2C_1C_2}} \qquad Q = \frac{\sqrt{R_1R_2C_1C_2}}{R_1C_1\,(1 - A_V) + C_2(R_1 + R_2)} \qquad (16\text{-}25)$$

Clearly, the five circuit parameters $R_1$, $R_2$, $C_1$, $C_2$, and $A_V$ provide more degrees of freedom than are needed to specify $\omega_o$ and $Q$. Often, to simplify fabrication, $C_1 = C_2 = C$ and $R_1 = R_2 = R$. Equations (16-24) and (16-25) for these conditions reduce to

$$H(s) = \frac{A_V}{R^2C^2s^2 + RCs(3 - A_V) + 1} \qquad (16\text{-}26)$$

$$\omega_o = \frac{1}{RC} \qquad Q = \frac{1}{3 - A_V} \qquad (16\text{-}27)$$

In Eq. (16-27), observe that making both resistors and both capacitors equal results in the time constant $RC$ establishing $\omega_o$ and the gain $A_V$ determining $Q$. Furthermore, note $A_V < 3$ for stability. If $A_V \geq 3$, then from Eq. (16-26), the $s$-coefficient is $\leq 0$ indicating $H(s)$ has right-half-plane poles.

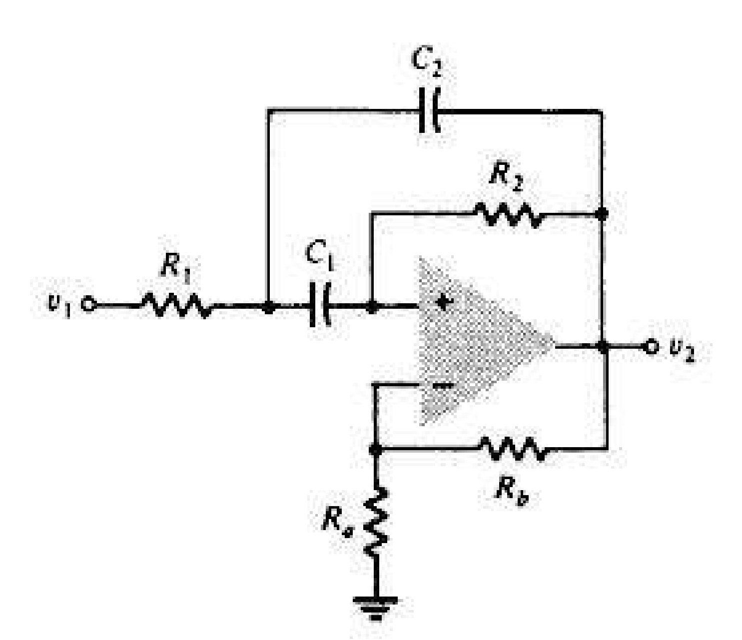



**FIGURE 16-35**
The Delyiannis band-pass circuit using both positive and negative feedback.

The band-pass sections in Figs. 16-34 and 16-35 are useful in realizing narrow-band circuits; that is, the passband is a fraction of the center frequency. The response of narrow-band circuits is similar to that obtained with a simple series or parallel resonant circuit having moderate $Q$. For this situation, the upper and lower 3-dB frequencies $f_2$ and $f_1$, respectively (Fig. 16-26*b*), are quite close so that the center frequency $f_o = \sqrt{f_1 f_2} \approx (f_1 + f_2)/2$ and the bandwidth $f_2 - f_1 = f_o/Q$. In some applications $f_2 - f_1 > f_o$ and broad-band band-pass circuits are required. This characteristic can be obtained by cascading low-pass and high-pass sections as shown in Fig. 16-36*a*. If $\omega_H > \omega_L$ (Fig. 16-36*b*), the band-pass response shown in Fig. 16-36*c* results. Note in Fig. 16-36*c* that attenuation at low frequencies ($\omega < \omega_L$) is provided by the high-pass network, whereas for $\omega > \omega_H$, attenuation results from the low-pass section. Both networks transmit the signal in the passband ($\omega_L \leq \omega \leq \omega_H$).

The cascade structure cannot be used in the narrow-band case because of component variations (sensitivity). If $\omega_L$ and $\omega_H$ are nearly equal, a small change in either (both) causes a significant error in the passband (Prob. 16-32).

**Band-Reject Sections** The parallel-channel configuration in Fig. 16-37*a* is used to realize a broad-band band-reject filter. If the low-pass and high-pass networks have frequency responses depicted in Fig. 16-36*b* and $\omega_H < \omega_L$, then Fig. 16-37*b* is the response of the circuit in Fig. 16-37*a*. Both the high-pass and low-pass sections provide attenuation in the stopband between $\omega_H$ and $\omega_L$. For $\omega < \omega_H$, transmission is through the low-pass section and for $\omega > \omega_L$, the signal is transmitted through the high-pass section.

Narrow-band band-reject circuits are often called *notch filters*. This can be discerned from the band-elimination entry in Table 16-2 for which $H(j\omega_r) = 0$ when $s = j\omega_r$. Note that $H(j\omega) \neq 0$ for all $\omega \neq \omega_r$ as displayed in the frequency response shown in Fig. 16-38. The circuit in Fig. 16-39 is commonly used to obtain the response shown in Fig. 16-38. The passive elements form a twin-tee network which provides the $j$-axis zeros. As depicted in Fig. 16-39, selection of $R_1 = R_2 = R$, $C_1 = C_2 = C$, $R_3 = R/2$, and $C_3 = 2C$ results in a transfer function (with $Y = 0$)

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

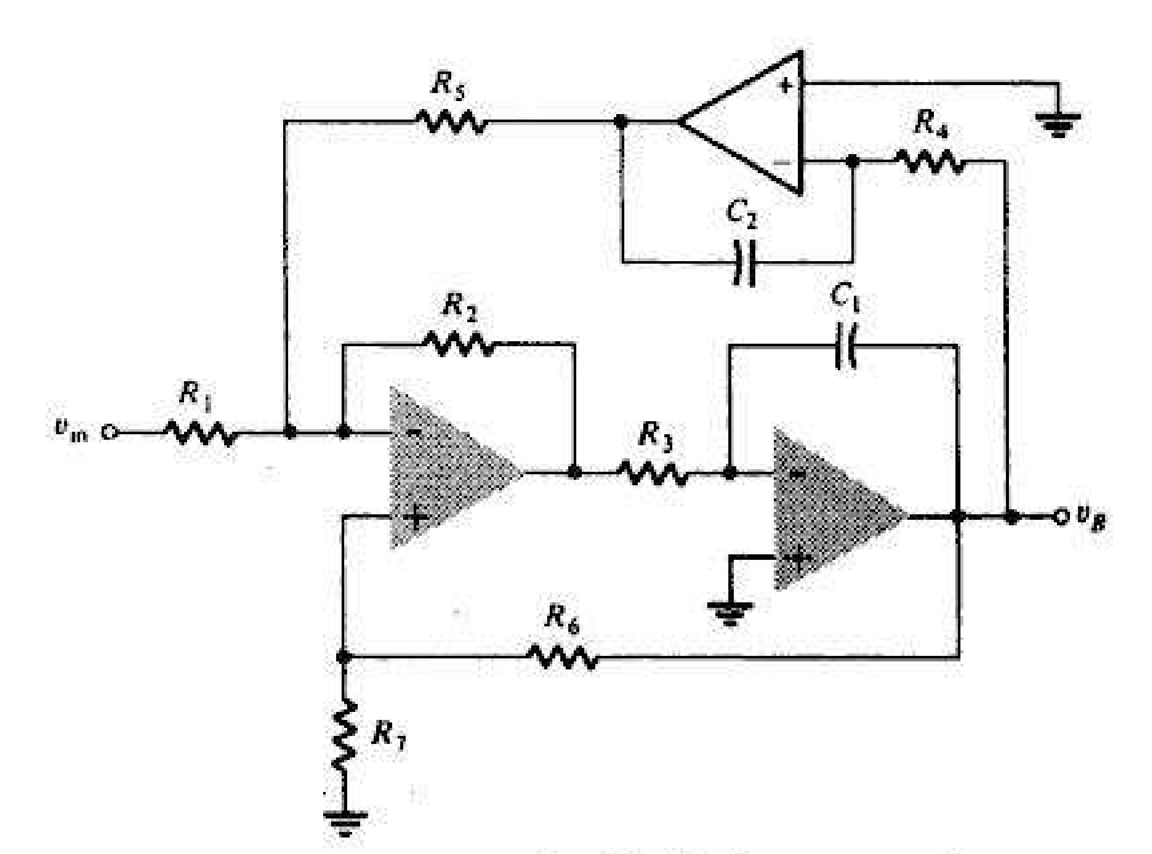



**FIGURE 16-45**
The circuit in Fig. 16-44 redrawn to show $v_B$ as the output. Note that the integrator formed by $R_4$, $C_2$, and the Op-Amp are part of the feedback loop.

Qualitatively, we can demonstrate that $v_C$ is the low-pass output as follows. Consider the circuit in Fig. 16-44 as a feedback amplifier in which $R_5$ and $R_1$ form a resistive feedback loop around the amplifier contained in the blue rectangle. The gain without feedback $A(s) = A_o/D(s)$, where $D(s)$ has two roots, one attributed to each of the integrators in the amplifier. The return ratio $T = -\beta A(s) = -\beta A_o/D(s)$ contains the same two poles as $A(s)$ since $\beta$ is real because the feedback network is resistive. Hence, $A_F(s) = A(s)/[1 + T(s)] = A_o/[D(s) - \beta A_o]$. Clearly, $A_F$ has two poles and no finite zeros and is a low-pass biquadratic function.

The circuit in Fig. 16-45 is the same as that in Fig. 16-44 redrawn. In this feedback amplifier, one integrator is part of the feedback network (shown in color). Thus $\beta$ is not real but contains a pole at $s = 0$ due to the integrator, that is, $\beta = K/s$, where $K$ is a real constant. The amplifier includes only one integrator and thus has only one pole $A(s) = A_o'/(1 + s/p_1)$. Again, use of Eq. (12-5) yields

$$A_F(s) = \frac{A_o'/(1 + s/p_1)}{1 + [A_o'/(1 + s/p_1)] \times (K/s)} = \frac{A_o' s}{D(s)}$$

where $D(s)$ is a quadratic. Comparison of $A_F(s)$ with the band-pass entry in Table 16-2 shows that they are equivalent.

In similar fashion, we can show that $v_A$ is the high-pass output as both integrators are part of the feedback network and that the basic amplifier, consisting only of the inverting Op-Amp stage, is frequency-independent.

The various transfer functions can be expressed as (Prob. 16-44)

$$H_L = \frac{V_C}{V_{in}} = \frac{-R_5/R_1}{D(s)} \qquad H_B = \frac{V_B}{V_{in}} = \frac{R_4R_5C_2s/R_1}{D(s)} \qquad (16\text{-}38)$$

$$H_H = \frac{V_A}{V_{in}} = \frac{-R_3R_4R_5C_1C_2s^2/R_1}{D(s)}$$

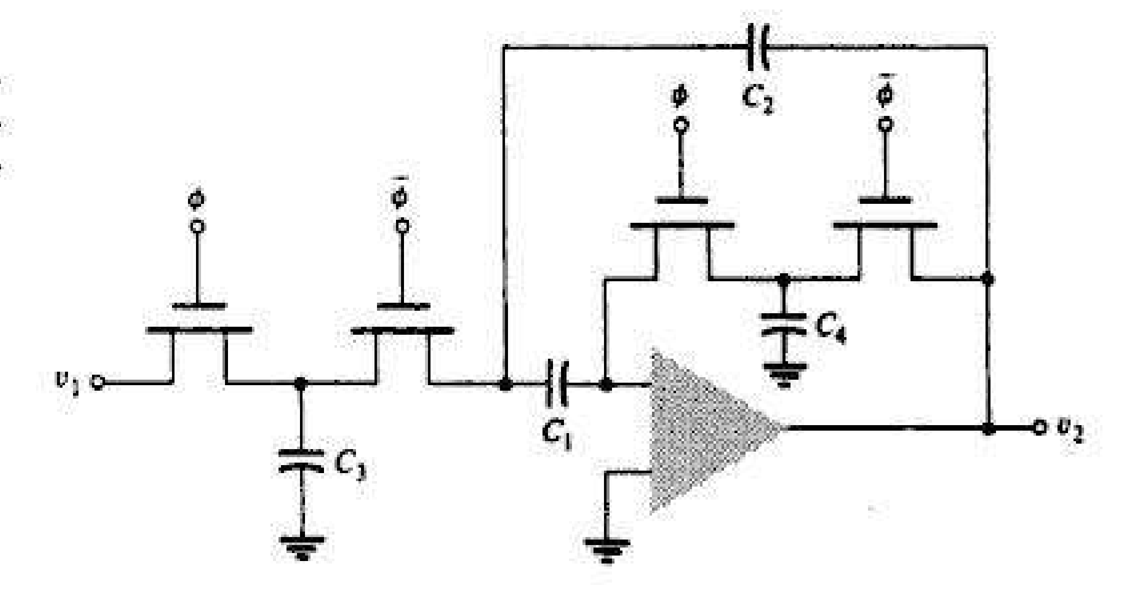



**FIGURE 16-51**
Switched-capacitor realization of the bandpass circuit in Fig. 16-34*b*.

## 16-13 LOGARITHMIC AND EXPONENTIAL AMPLIFIERS

In Fig. 16-52 an Op-Amp is shown with the feedback resistor $R_2'$ replaced by the diode $D1$. This amplifier is used when it is desired to have the output voltage proportional to the logarithm of the input voltage.

From Eq. (2-3) the volt-ampere diode characteristic is

$$i_f = I_S(\epsilon^{v_f/\eta V_T} - 1) \approx I_S\epsilon^{v_f/\eta V_T}$$

provided $v_f/\eta V_T \gg 1$. Hence

$$v_f = \eta V_T(\ln i_f - \ln I_S) \qquad (16\text{-}48)$$

Since $i_s = v_s/R$ due to the virtual ground at the amplifier input, then

$$v_o = -v_f = -\eta V_T\left(\ln \frac{v_s}{R} - \ln I_S\right) \qquad (16\text{-}49)$$

**Logarithmic Amplifier Using Matched Transistors** We note from Eq. (16-49) that $v_o$ is temperature-dependent as a result of the scale factor $\eta V_T$ and the saturation current $I_S$. The factor $\eta$, whose value normally depends on the diode current, can be eliminated by replacing the diode with a grounded-base transistor. Another important advantage of using a transistor in place of a diode is that the exponential relationship between current and voltage extends over a much wider voltage range for a transistor than a diode. By augmenting Fig. 16-52

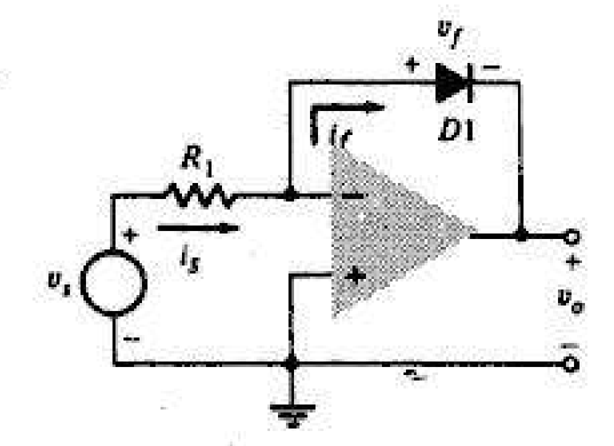



**FIGURE 16-52**
An elementary logarithmic amplifier.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

16-6 Draw the block diagram from which to obtain the gating signals for time-division multiplexing.

16-7 (*a*) Draw a schematic diagram of a D/A converter. Use resistance values whose ratios are multiples of 2.
(*b*) Explain the operation of the converter.

16-8 Indicate two possible implementations for the digitally controlled switch of a D/A converter.

16-9 Repeat Rev. 16-7 for a ladder network whose resistances have one of two values, $R$ or $2R$.

16-10 Explain how a DAC may function as a programmable attenuator for an analog signal.

16-11 (*a*) Draw the block diagram for a counting A/D converter.
(*b*) Explain the operation for this system.

16-12 Repeat Rev. 16-11 for a servo ADC.

16-13 Repeat Rev. 16-11 for a 2-bit parallel-comparator A/D converter.

16-14 Explain by means of an asymptotic Bode diagram why a practical integrator deviates from an ideal one at both high and low frequencies.

16-15 Sketch the circuit of a differential integrator and explain its operation.

16-16 Show how an integrator can be modified to become a differentiator.

16-17 Sketch the ideal frequency-response characteristics of low-pass, high-pass, band-pass, and band-reject systems.

16-18 (*a*) Write the transfer function for a general biquadratic function.
(*b*) What coefficients in part *a* must be zero in order to obtain a low-pass characteristic?
(*c*) Repeat part *b* for high-pass and band-pass characteristcs.
(*d*) Repeat part *b* for band-reject response.

16-19 Define by means of a diagram (*a*) the passband, (*b*) the stopband, (*c*) the transition band, (*d*) passband ripple.

16-20 What differences exist in the frequency response of Butterworth and Chebyshev filters (low-pass) of the same degree?

16-21 Draw the circuit diagram of a low-pass Sallen and Key positive-feedback section.

16-22 How can the circuit in Rev. 16-21 be modified to become a high-pass section?

16-23 Repeat Rev. 16-21 for a negative-feedback section.

16-24 Draw a block diagram of a broad-band band-pass filter and explain its operation.

16-25 Repeat Rev. 16-24 for a band-reject filter.

16-26 Draw the circuit diagram of a band-pass section using an ideal Op-Amp.

16-27 (*a*) What is meant by a notch network?
(*b*) Define high-pass and low-pass notches.

16-28 (*a*) What is meant by an all-pass network?
(*b*) Of what use is such a network?

By use of Eqs. (17-1), it follows that

$$I_{\text{rms}} = \left(\frac{1}{2\pi}\int_o^{\pi} I_m^2 \sin^2\alpha \, d\alpha\right)^{1/2} = \frac{I_m}{2} \tag{17-8}$$

The rms output voltage is given by $I_mR_L/2$.

Applying Eq. (17-7) to the *sinusoidal input voltage*, we obtain

$$V_{rms} = \frac{V_m}{\sqrt{2}} \tag{17-9}$$

**Regulation**

The variation of dc output voltage as a function of dc load current is called *regulation*. The percentage regulation is defined as

$$\% \text{ regulation} \equiv \frac{V_{\text{no load}} - V_{\text{load}}}{V_{\text{load}}} \times 100\% \tag{17-10}$$

where *no load* refers to zero current and *load* indicates the normal load current. For an ideal power supply the output voltage is independent of the load (the output current) and the percentage regulation is zero.

The variation of $V_{\text{dc}}$ with $I_{\text{dc}}$ for the half-wave rectifier is obtained as follows: From Eqs (17-4) and (17-2),

$$I_{\text{dc}} = \frac{I_m}{\pi} = \frac{V_m/\pi}{R_f + R_L} \tag{17-11}$$

Solving Eq. (17-11) for $V_{\text{dc}} = I_{\text{dc}}R_L$, we obtain

$$V_{\text{dc}} = \frac{V_m}{\pi} - I_{\text{dc}}R_f \tag{17-12}$$

This result is consistent with the circuit model given in Fig. 17-4 for the dc voltage and current. Note that the rectifier circuit functions as if it were a constant (open-circuit) voltage source $V = V_m/\pi$ in series with an effective internal resistance (the *output resistance*) $R_o = R_f$. This model shows that $V_{\text{dc}}$ equals $V_m/\pi$ at no load and that the dc voltage decreases linearly with an increase in dc output current. In practice, the resistance $R_s$ of the transformer secondary is in series with the diode, and in Eq. (17-12) $R_s$ should be added

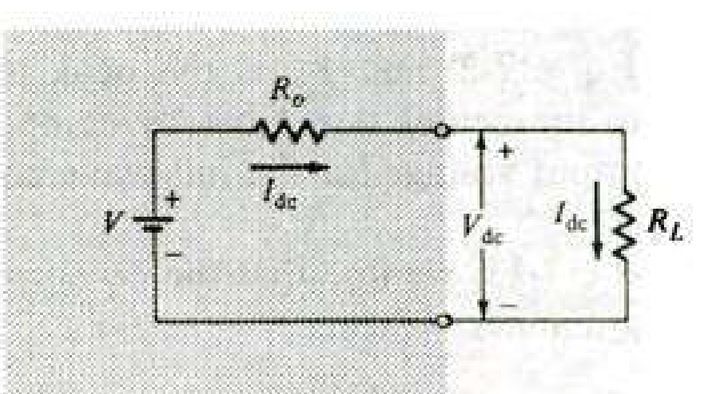



**FIGURE 17-4**
The Thévenin equivalent of a power supply used to determine the load voltage and current.

voltage varies with the temperature, particularly because semiconductor devices are used.

The Zener diode can be used as a simple regulator in the circuit shown in Fig. 2-32. This circuit, as described in Sec. 2-11, is limited by the current (and power) capability of the Zener diode used. Typically, the Zener diode must be able to handle a current that is greater than that supplied to the lóad.

The feedback circuit shown in Fig. 17-15 is used to overcome the three shortcomings described previously and the current limitation of the Zener diode. Such a system is called a *regulated power supply*. From Fig. 17-15 we see that the regulated power supply represents a case of series-shunt (voltage-series feedback). If we assume that the voltage gain of the emitter follower $Q1$ ($Q1$ is also called the *pass transistor* or *element*) is approximately unity, then $V'_O \approx V_O$ and

$$V'_O = A_V V_i = A_V(\beta V_O - V_R) \approx V_O \tag{17-23}$$

where

$$\beta \equiv \frac{R_2}{R_1 + R_2} \tag{17-24}$$

From Eq. (17-23) it follows that

$$V_O = V_R \frac{A_V}{1 + \beta A_V} \tag{17-24}$$

If $\beta A_V \gg 1$, $V_O \approx V_R/\beta$; note, however, that $V_O$ must be less than the unregulated voltage source $V_{dc}$.

The output voltage $V_O$ can be changed by varying $\beta$, by changing the fraction of $V_O$ that is fed back. The emitter follower $Q1$ is used to provide current gain, because the current delivered by Op-Amp $A_V$ usually is not sufficient. Also, the pass element must absorb the difference between the unregulated input

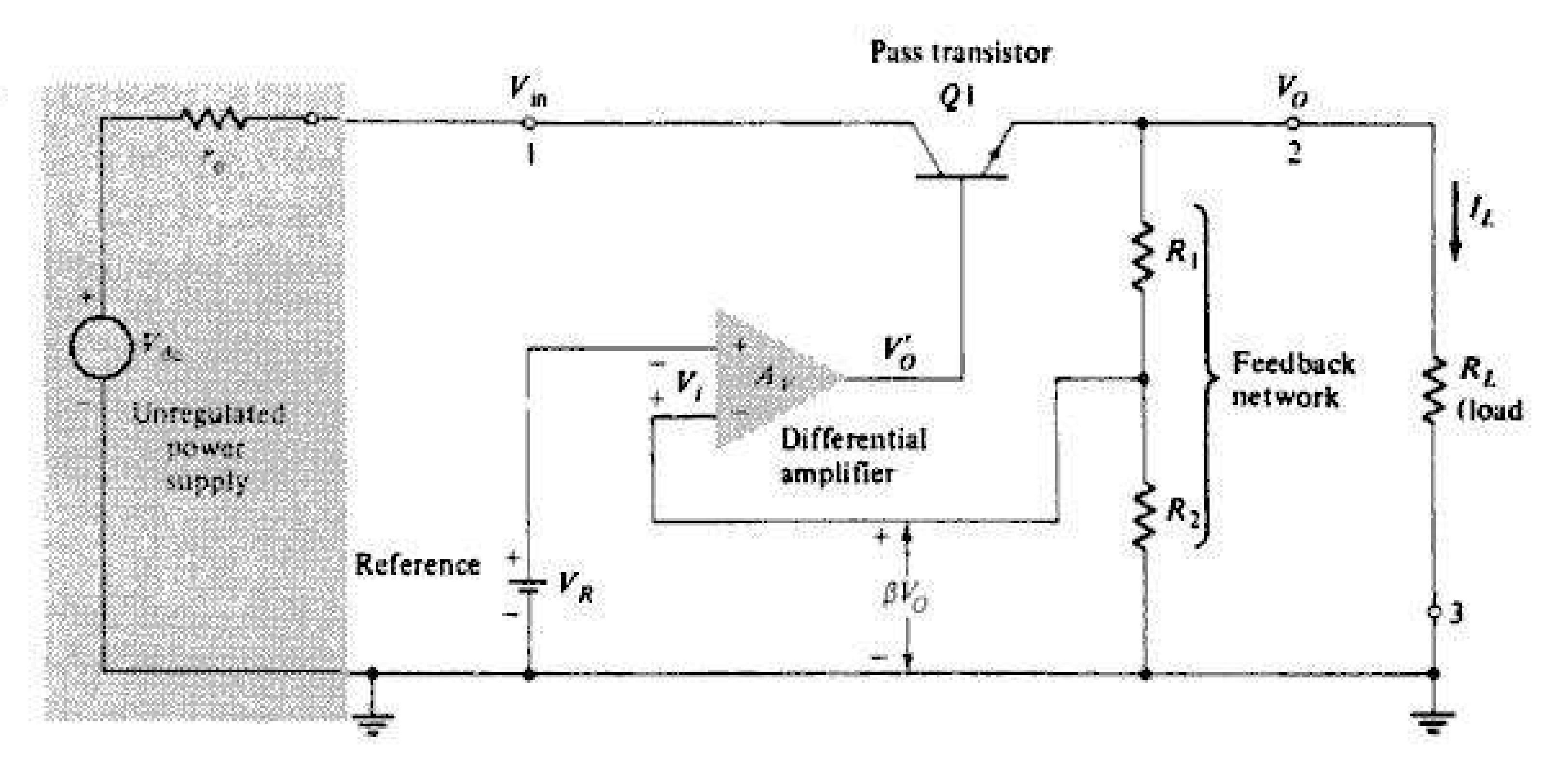



**FIGURE 17-15**
A regulated power-supply system.

**FIGURE 17-24**
The block diagram of the system used to generate the waveforms $v_{A1}$ and $v_{A2}$ in Fig. 17-23 from the pulse-width modulator output $v_A$. The power switch transistors $Q1$ and $Q2$ and the base-drive transistors $Q3$ and $Q4$ are also shown.

for the pulse-width modulator. This PWM waveshape $v_A$ is given in Fig. 17-25*b*. The duty cycle $\delta$ of $v_A$ is $T_1/(T_1 + T_2)$. The FLIP-FLOP is used as a divide-by-2 circuit, whose input is $v_{osc}$, and the two complementary FLIP-FLOP outputs $v_Q$ and $v_{\bar{Q}}$ are shown in Fig. 17-25*c* and *d*, respectively. The inputs to AND gate $A1$ ($A2$) are $v_Q$ ($v_{\bar{Q}}$) and $v_A$, and the outputs $v_{A1}$ and $v_{A2}$ are drawn in Fig. 17-25*b* and *f* respectively. These are the waveforms used in Figs. 17-25*b* and 17-25*c*.

The switch $SW1$ ($SW2$) in Fig. 17-22 is replaced by the power transistor $Q1$ ($Q2$) in Fig. 17-24. The base currents for $Q1$ and $Q2$ are supplied by transistors $Q3$ and $Q4$, which are driven by waveforms $v_{A1}$ and $v_{A2}$ respectively. The complexity of the switching regulator system would preclude its usefulness were it not for the increased level of sophistication attainable in modern microelectronics. The Silicon General SG1524 package is such an example.

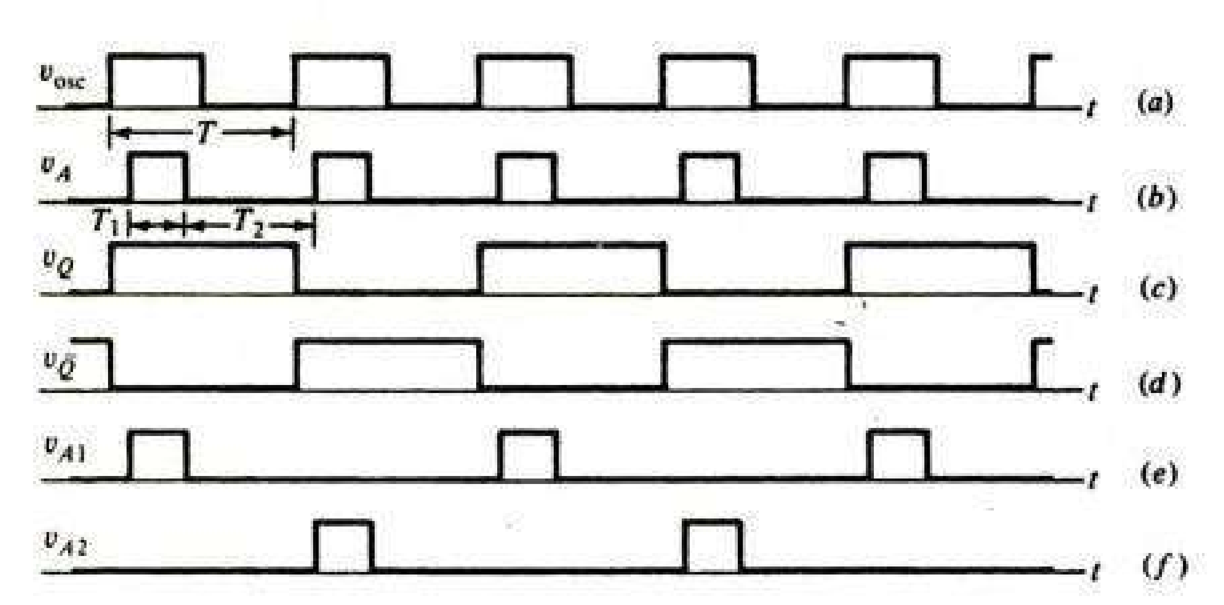



**FIGURE 17-25**
The waveforms shown in Fig. 17-24. (All amplitudes are drawn equal for simplicity.)

necessary to express the dynamic transfer curve with respect to the $Q$ point by a power series of the form

$$i_c = G_1 i_b + G_2 i_b^2 + G_3 i_b^3 + G_4 i_b^4 + \cdots \qquad (17\text{-}46)$$

If we assume that the input wave is a simple cosine function of time, of the form in Eq. (17-38), the output current will be given by

$$i_C = I_C + B_0 + B_1 \cos \omega t + B_2 \cos 2\omega t + B_3 \cos 3\omega t + \cdots \qquad (17\text{-}47)$$

This equation results when Eq. (17-38) is inserted in Eq. (17-46) and the proper trigonometric transformations are made.

Note that now a third harmonic and higher-order harmonics are present. The Fourier coefficients $B_0, B_1, B_2, B_3, \ldots$ may be obtained by an extension of the foregoing procedure used with Eq. (17-47) instead of Eq. (17-39).

The harmonic distortion is defined as

$$D_2 \equiv \frac{|B_2|}{|B_1|} \qquad D_3 \equiv \frac{|B_3|}{|B_1|} \qquad D_4 \equiv \frac{|B_4|}{|B_1|} \qquad (17\text{-}48)$$

where $D_s$ ($s = 2, 3, 4, \ldots$) represents the distortion of the $s$th harmonic.

**Power Output**

If the distortion is not negligible, the power delivered at the fundamental frequency is

$$P_1 = \frac{B_1^2 R_L}{2} \qquad (17\text{-}49)$$

However, the total power output is

$$P = (B_1^2 + B_2^2 + B_3^2 + \cdots)\frac{R_L}{2} = (1 + D_2^2 + D_3^2 + \cdots)P_1$$

or

$$P = (1 + D^2)P_1 \qquad (17\text{-}50)$$

where *the total harmonic distortion* (THD), or *distortion factor*, is defined as

$$D \equiv \sqrt{D_2^2 + D_3^2 + D_4^2 + \cdots} \qquad (17\text{-}51)$$

If the total distortion is 10 percent of the fundamental, then

$$P = [1 + (0.1)^2]P_1 = 1.01P_1$$

The total power output is only 1 percent higher than the fundamental power when the distortion is 10 percent. Hence little error is made in using only the fundamental term $P_1$ in calculating the power output.

In passing, it should be noted that the total harmonic distortion is not necessarily indicative of the discomfort to someone listening to music. Usually, the same amount of distortion is more irritating, the higher the order of the harmonic frequency.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

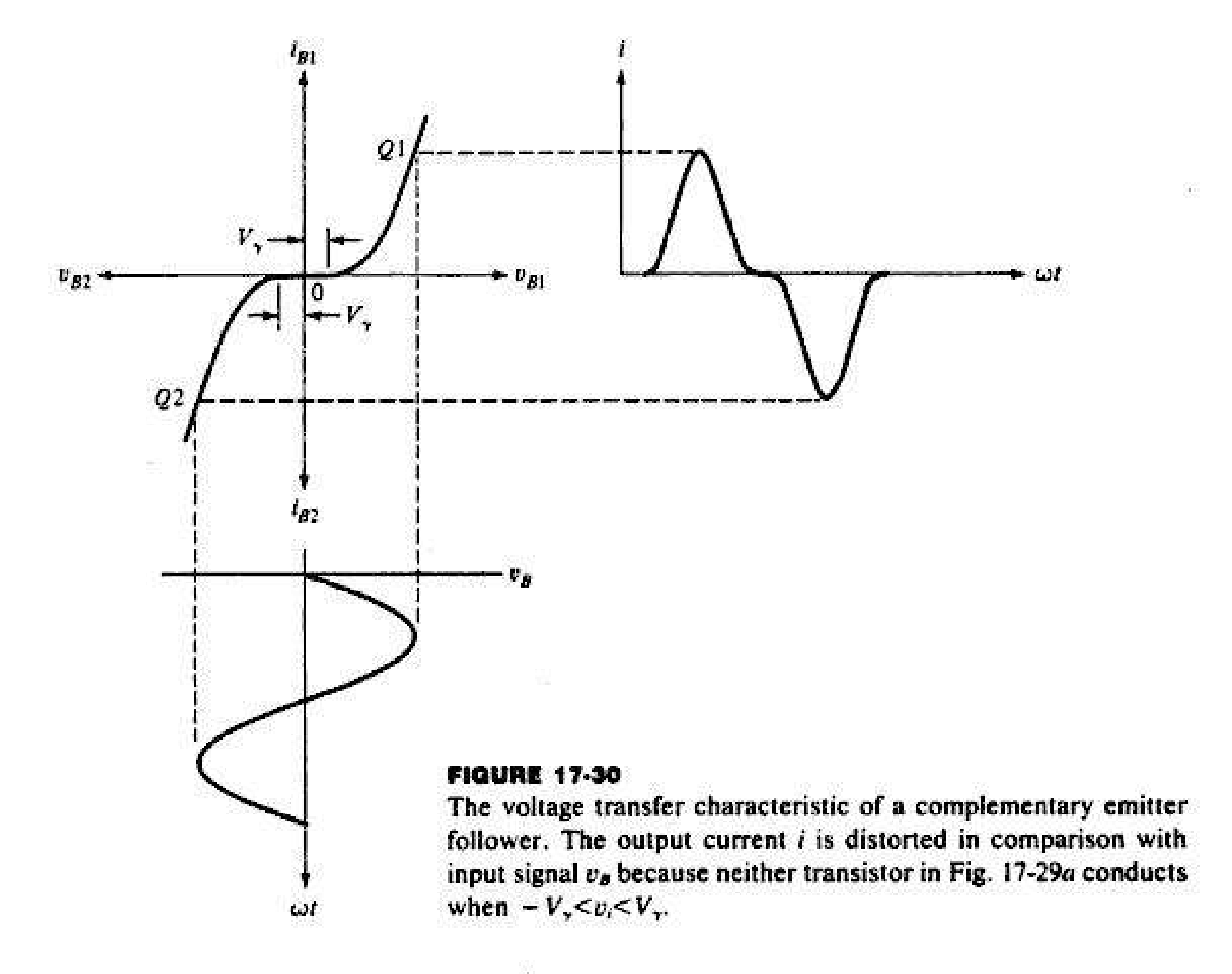



**FIGURE 17-30**
The voltage transfer characteristic of a complementary emitter follower. The output current $i$ is distorted in comparison with input signal $v_B$ because neither transistor in Fig. 17-29*a* conducts when $-V_\gamma < v_i < V_\gamma$.

## 17-14 CLASS AB OPERATION

In addition to the distortion introduced by not using matched transistors and that due to the nonlinearity of the collector characteristics, there is one more source of distortion, that caused by nonlinearity of the input characteristic. As pointed out in Sec. 3-3 and Fig. 3-9, no appreciable base current flows until the emitter junction is forward-biased by at least the cut-in voltage $V_\gamma$, which is 0.5 V for silicon. Under these circumstances a sinusoidal base-voltage excitation will not result in a sinusoidal output current. Although already mentioned briefly in Sec. 14-5, the significance of the nonlinear input characteristic merits further discussion.

The distortion caused by this nonlinear curve is indicated in Fig. 17-30. The $i_B$-$v_B$ curve for each transistor is drawn, and the construction used to obtain the output current (assumed proportional to the base current) is shown. In the region of small currents (for $v_B < V_\gamma$) the output is much smaller than it would be if the response were linear. This effect is called *crossover distortion*. Such distortion would not occur if the driver were a true current generator, in other words, if the base current (rather than the base voltage) were sinusoidal.

To minimize crossover distortion, the transistors must operate in a class AB mode, where a small standby current flows at zero excitation. For example, in the circuit of Fig. 14-14 the difference between the base voltages of the two transistors is adjusted to be approximately equal to $2V_\gamma$. Class AB operation results in less distortion than class B, but the price which must be paid for this

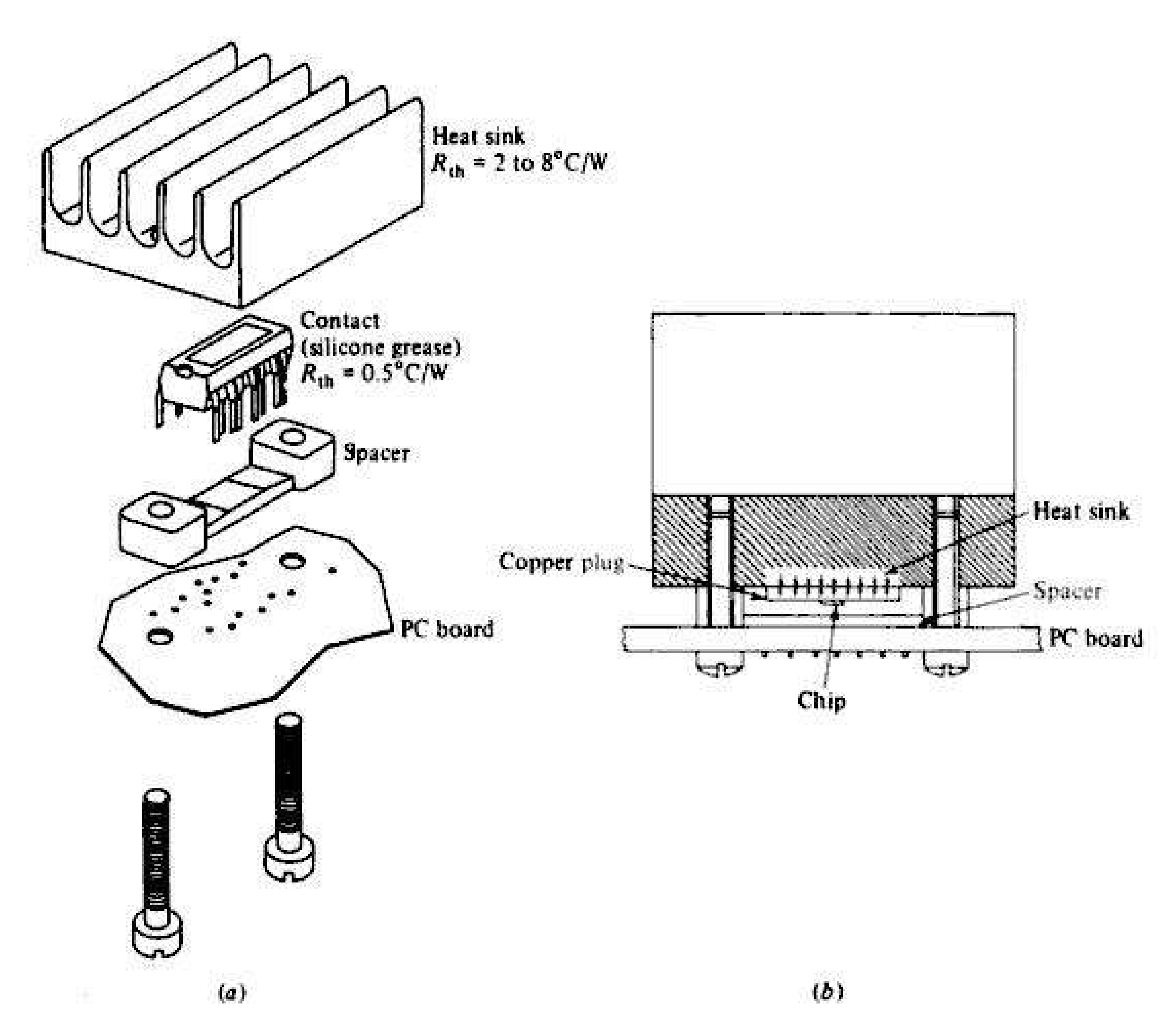



**FIGURE 17-34**
Heat sink mounting of TDA2020. A number of heat sinks are available with thermal resistances in the range 2 to 8°C/W. (*b*) Cross section of assembled system. (*Courtesy of SGS/AETA*)

with a relatively large heat-radiating surface to which the transistor case is attached. Figure 17-34 depicts the mounting system for the TDA2020 chip in Fig. 17-32.

***Example 17-1*** In the foregoing discussion we observed that the 20-W TDA2020 amplifier must dissipate 15.1 W of internal power. The ambient temperature is $T_A = 30°C$. If the maximum allowable junction temperature is $T_{J(\max)} = 150°C$ and if $\theta_{JC} = 3°C/W$, what is the maximum heat-sink thermal resistance $\theta_{SA}$ (sink-to-ambient) that can be tolerated?

***Solution*** Using the electrical analog of Eq. (17-67), we obtain the series circuit model given in Fig. 17-35 for the power flow.

$$\begin{aligned} T_J &= \Delta T_{JC} + \Delta T_{CS} + \Delta T_{SA} + T_A \\ &= P_D(\theta_{JC} + \theta_{CS} + \theta_{SA}) + T_A \end{aligned} \qquad (17\text{-}68)$$

17-17 List three disadvantages of pass regulators which may be overcome with a switching regulator.

17-18 (*a*) Draw the basic switching-regulator topology.
(*b*) Explain how the dc output voltage is determined by this feedback system.

17-19 Explain why a switching regulator is capable of very high conversion efficiency.

17-20 (*a*) Draw the power switch of a switching regulator as an SPDT switch. How is the switch controlled, and what is the switch output waveform?
(*b*) Verify that the switch in part (*a*) is equivalent to an SPST switch in series with a diode to ground. Why is the diode referred to as a flyback diode?

17-21 (*a*) Indicate the SPDT power switch of the basic regulator as a combination of three transistors and a diode.
(*b*) Explain the function of each transistor and the diode.

17-22 (*a*) Draw the power components in a switching regulator for which $V_O$ is positive and greater than $V_{in}$.
(*b*) Verify that for this configuration $V_O > V_{in}$.
(*c*) What determines the numerical value of $V_O$?

17-23 (*a*) Repeat Rev. 17-22 for a regulator for which $V_O$ is negative.
(*b*) Give an argument to justify that $V_O < 0$.

17-24 (*a*) Draw the power switch for a push-pull transformer-coupled switching regulator.
(*b*) Indicate the pulse-width-modulator waveform $v_A$, and also the waveforms $v_{A1}$ and $v_{A2}$ controlling the two SPST switches in series with the transformer primaries.
(*c*) Sketch the transformer secondary waveforms.
(*d*) Draw the waveform from the output switch (the input voltage to the filter).

17-25 (*a*) Draw in block-diagram form the system for obtaining the waveforms $v_{A1}$ and $v_{A2}$ in Rev. 17-24, part *b*.
(*b*) Explain the operation of the system with the aid of a waveform chart.
(*c*) Show the switches controlled by $v_{A1}$ and $v_{A2}$ simulated by transistors.

17-26 List all the low-power control circuits which are fabricated on a single IC chip and used with a switching regulator.

17-27 Derive an expression for the output power of a class A large-signal amplifier in terms of $V_{max}$, $V_{min}$, $I_{max}$, and $I_{min}$.

17-28 Discuss how rectification may take place in a power amplifier.

17-29 Define intermodulation distortion.

17-30 Define total harmonic distortion.

17-31 Define a (*a*) class A, (*b*) class B, and (*c*) class AB amplifier.

17-32 (*a*) Define the conversion efficiency $\eta$ of a power stage.
(*b*) Derive a simple expression for $\eta$ for a class A amplifier.
(*c*) What is the theoretical maximum efficiency for a class A amplifier?

17-33 (*a*) Draw the circuit of a class B power stage.
(*b*) For a sinusoidal input, what is the output waveform?

# *Appendix B*
# SEMICONDUCTOR MANUFACTURERS AND DEVICE SPECIFICATIONS

**B-1 ELECTRONIC DEVICE MANUFACTURERS** Databooks and applications information may be obtained from the following semiconductor companies:

**Advanced Micro Devices** 901 Thompson Pl., Sunnyvale, CA 94086
**American Microsystems Inc.** 3800 Homestead Road, Santa Clara, CA 95051
**Analog Devices** 2 Technology Way, Norwood, MA 02062
**Burr-Brown Research Corp.** 6730 S. Tucson Blvd., Tucson, AZ 85734
**Fairchild Semiconductor** 464 Ellis St., Mountain View, CA 94042
**Ferranti Electric** E. Bethpage Rd., Plainview, NY 11803
**General Electric Co.** Schenectady, NY 13201
**General Instrument Corp.** 600 West John St., Hicksville, NY 11802
**Harris Semiconductor** Box 833, Melbourne, FL 32901
**Hitachi America, Ltd.** 111 E. Wacker Dr., Chicago, IL 60601
**Imsai** 14860 Wicks Blvd., San Leandro, CA 94577
**Intel Corp.** 3065 Bowers Ave., Santa Clara, CA 95051
**Intersil Inc.** 10900 N. Tantau Ave., Cupertino, CA 95014
**ITT Semiconductors** 74 Commerce Way, Woburn, MA 01801
**Monolithic Memories, Inc.** 1165 E. Argues Ave., Sunnyvale, CA 94086
**Mostek Corp.** 1215 W. Crosby Rd., Carollton, TX 75006
**Motorola Semiconductor Products** Box 20912, Phoenix, AZ 85036
**National Semiconductor, Inc.** 2900 Semiconductor Dr., Santa Clara, CA 95051
**Plessey Semiconductors** 1674 McGraw Ave., Santa Ana, CA 92705
**Raytheon Semiconductor** 350 Ellis St., Mountain View, CA 94042
**RCA Solid State Division** Box 3200, Somerville, NJ 08876
**SGS/ATES Semiconductor Corp.** 796 Massasoit Street, Waltham, MA 03254
**Signetics Corp.** 811 E. Argues Ave., Sunnyvale, CA 94086
**Silicon General** 73826 Bolsoo Ave., Westminster, CA 92683
**Siliconix, Inc.** 2201 Laurelwood Road, Santa Clara, CA 95054
**Stewart-Warner Microcircuits** 730 E. Evelyn Ave., Sunnyvale, CA 94086

## B-7 SPECIFICATIONS FOR LM741 OPERATIONAL AMPLIFIER (Courtesy of National Semiconductor, Inc.)

A high-performance monolithic Op-Amp intended for a wide range of analog applications. It is short-circuit protected and requires no external components for frequency compensation. The LM741C is identical to the LM741, except that the performance of the former is guaranteed over a 0 to 70°C temperature range, instead of −55 to 125°C.

**TABLE B7-1 Absolute Maximum Ratings**

| | |
|---|---|
| Supply voltage LM741 | ±22 V |
| LM741C | ±18 V |
| Power dissipation* | 500 mW |
| Differential input voltage | ±30 V |
| Input voltage† | ±15 V |
| Output short-circuit duration | Indefinite |
| Storage temperature range | −65 to 150°C |
| Lead temperature (soldering, 10 s) | 300°C |

*The maximum junction temperature of the LM741 is 150°C, while that of the LM741C is 100°C. For operating at elevated temperatures, devices in the TO-5 package must be derated based on a thermal resistance of 150°C/W, junction to case.

†For supply voltages less than ±15 V, the absolute maximum input voltage is equal to the supply voltage.

**TABLE B7-2 Electrical Characteristics***

| | | *LM741* | | | *LM741C* | | | |
|---|---|---|---|---|---|---|---|---|
| *Parameter* | *Conditions* | *Min* | *Typ* | *Max* | *Min* | *Typ* | *Max* | *Unit* |
| Input offset voltage | $T_A = 25°C$, $R_s \leq 10$ kΩ | | 1.0 | 5.0 | | 1.0 | 6.0 | mV |
| Input offset current | $T_A = 25°C$ | | 30 | 200 | | 30 | 200 | nA |
| Input bias current | $T_A = 25°C$ | | 200 | 500 | | 200 | 500 | nA |
| Input resistance | $T_A = 25°C$ | 0.3 | 1.0 | | 0.3 | 1.0 | | MΩ |
| Supply current | $T_A = 25°C$, $V_s = \pm 15$ V | | 1.7 | 2.8 | | 1.7 | 2.8 | mA |
| Large-signal voltage gain | $T_A = 25°C$, $V_s = \pm 15$ V<br>$V_{OUT} = \pm 10$ V, $R_L \geq 2$ kΩ | 50 | 160 | | 25 | 160 | | V/mV |
| Input offset voltage | $R_s < 10$ kΩ | | | 6.0 | | | 7.5 | mV |
| Input offset current | | | | 500 | | | 300 | nA |
| Input bias current | | | | 1.5 | | | 0.8 | μA |
| Large-signal voltage gain | $V_s = \pm 15$ V,<br>$V_{OUT} = \pm 10$ V<br>$R_L \geq 2$ kΩ | 25 | | | 15 | | | V/mV |
| Output voltage swing | $V_s = \pm 15$ V, $R_L = 10$ kΩ<br>$R_L = 2$ kΩ | ±12<br>±10 | ±14<br>±13 | | ±12<br>±10 | ±14<br>±13 | | V<br>V |
| Input voltage range | $V_s = \pm 15$ V | ±12 | | | ±12 | | | V |
| Common-mode rejection ratio | $R_s \leq 10$ kΩ | 60 | 90 | | 70 | 90 | | dB |
| Supply voltage rejection ratio | $R_s \leq 10$ kΩ | 77 | 96 | | 77 | 96 | | dB |

*These specifications apply for $V_s = \pm 15$ V and $-55°C \leq T_A \leq 125°C$, unless otherwise specified. With the LM741C, however, all specifications are limited to $0°C \leq T_A \leq 70°C$ and $V_s = \pm 15$ V.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

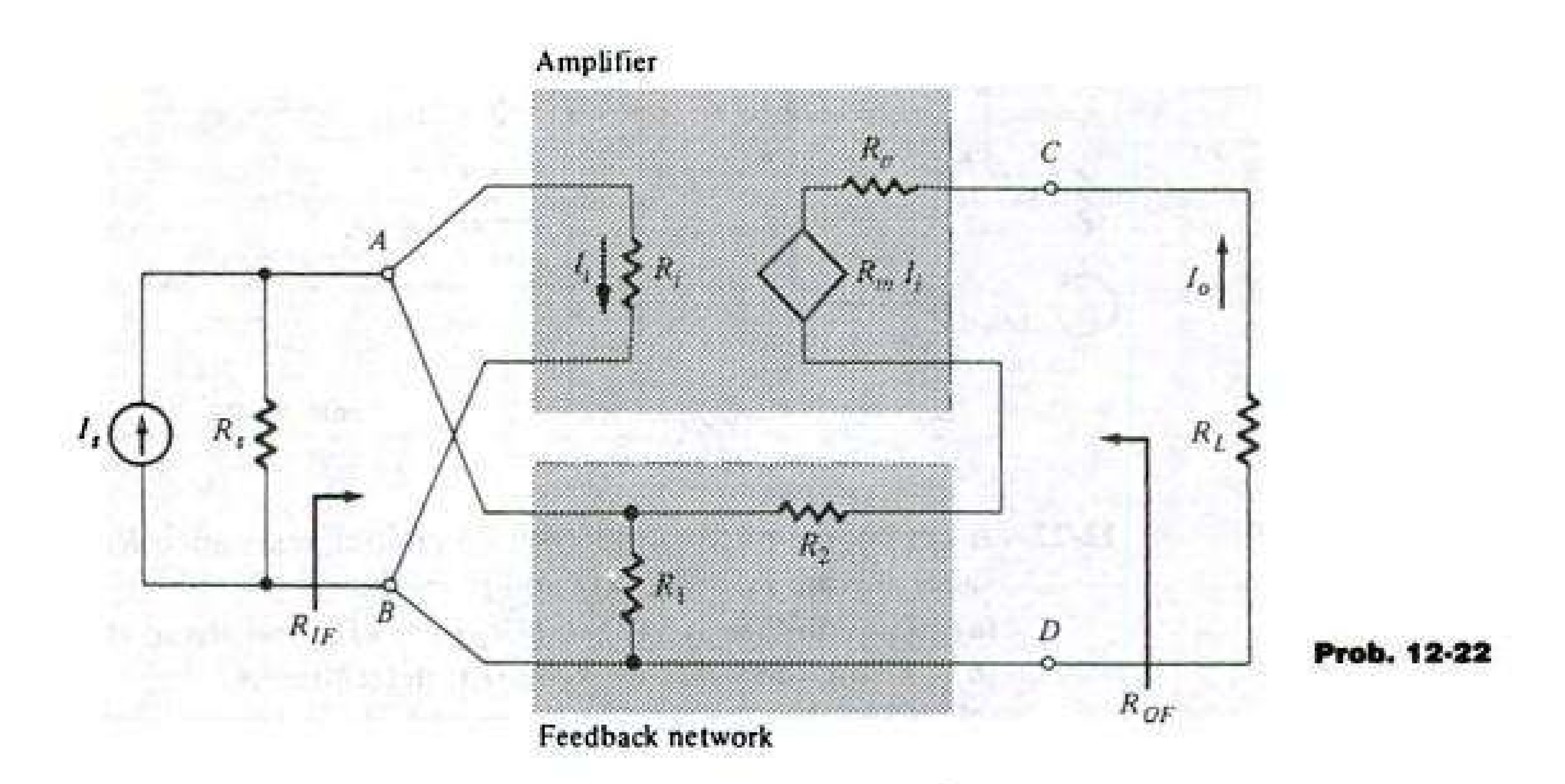
You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

**12-22** (*a*) Repeat Prob. 12-21 for the circuit shown. The element values are: $R_i$ = 5 kΩ, $R_o$ = 0.5 kΩ, $R_m$ = 100 kΩ, $R_2$ = 10 kΩ, $R_1$ = 1 kΩ, $R_s$ = 50 Ω, and $R_L$ = 2 kΩ.

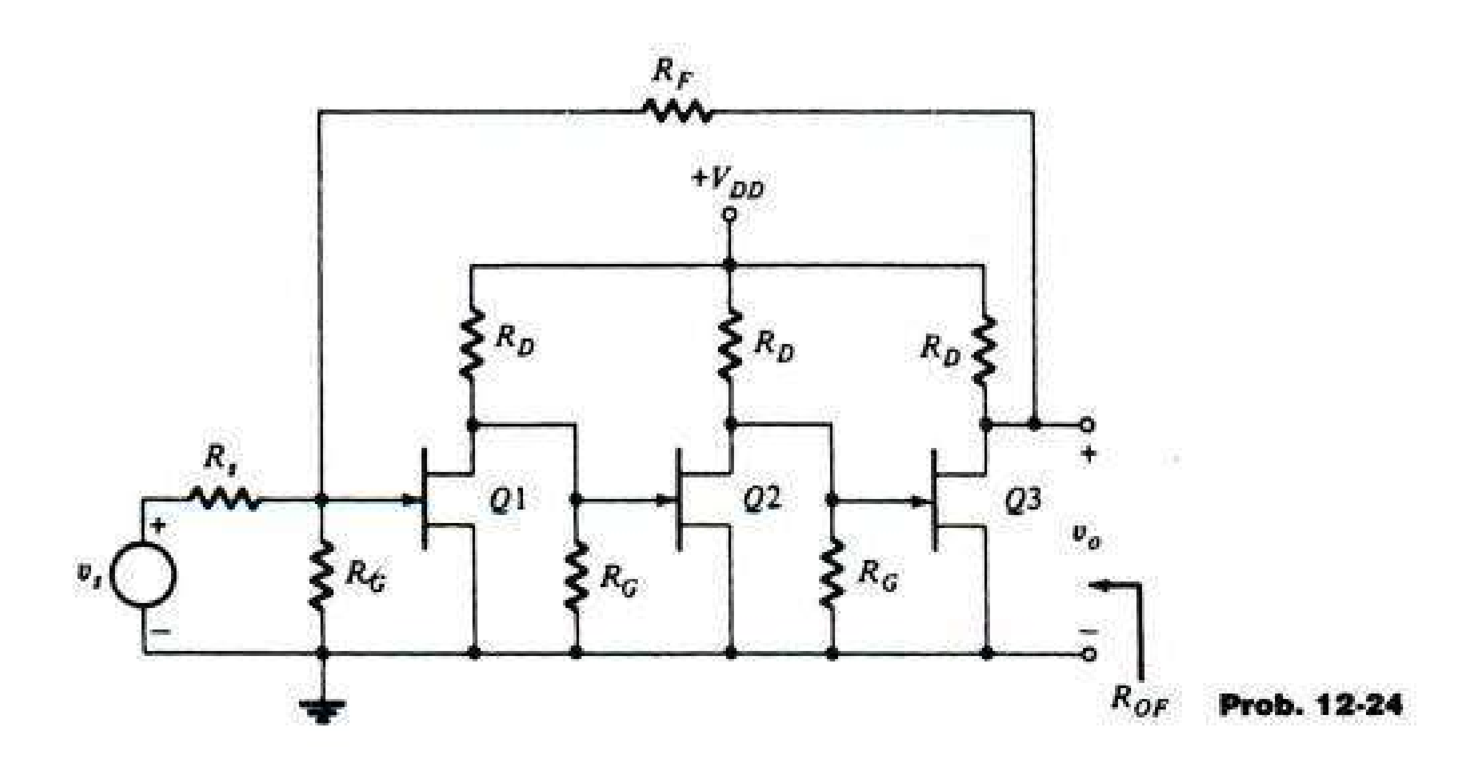
(*b*) What new value of $R_m$ is needed to make $R_{IF}$ = 50 Ω?




**Prob. 12-22**

**12-23** Verify Eqs. (12-41) and (12-42).

**12-24** The FETs in the circuit shown are identical and have $g_m$ = 2 m℧ and $r_d$ = 20 kΩ. The circuit parameters are $R_D$ = 12 kΩ, $R_G$ = 500 kΩ, $R_s$ = 50 Ω, and $R_F$ = 5 kΩ. Determine $A_F$ and $R_{OF}$.




**Prob. 12-24**

**12-25** Transistor *A*, biased at $I_{CQ}$ = 1.5 mA, is used in the circuit shown.

(*a*) Determine $A_F$ and *T*.

(*b*) Find $R_{IF}$ and $R_{OF}$.

(*c*) What value, if any, of *R* is needed to make $R_{OF}$ = 47 Ω?

**12-26** Verify Eq. (12-43).

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

## CHAPTER 5

**5-4** 37.8 Ω/square.

**5-5** (*a*) 2.5 mm; (*b*) 1 μm.

**5-7** 307 pF.

## CHAPTER 6

**6-2** 11011110, 100101110, 1111110000.

**6-3** 177, 550, 2052.

**6-4** DE, 12E, 6FO.

**6-6** $1.2 \leq R \leq 5.56$ kΩ.

**6-8** (*a*) $R_{OFF} \geq 120$ kΩ; (*b*) $R_{ON} \leq 208$ Ω.

**6-13** (*b*) $Y = \overline{ABC}$.

**6-15** (*b*) $Y = AB\overline{C}$.

**6-24** 22.

**6-25** (*a*) 52.6 kΩ.

**6-45** (*a*) $Y = (\overline{A + B})$.

**6-54** (*a*) 4.39; (*b*) 7.1 mW.

**6-56** (*a*) 1.7 V; (*b*) 27.

**6-62** 0.3 V, $i_b = 51.7$ μA, $i_c = 131.3$ μA.

**6-64** (*b*) 50; (*c*) 9.63 mW.

**6-67** (*c*) $V(0) = 0.2$ V, $V(1) = 2.746$ V; (*d*) 88.

**6-69** (*a*) 0; (*b*) 43.5 mA.

## CHAPTER 7

**7-17** (*a*) 145; (*b*) 145.

**7-18** (*a*) 80; (*b*) 60.

**7-22** (*b*) 9.

**7-25** 8, $X_0 = X_5 = \overline{D}$, $X_1 = X_4 = D$, $X_2 = X_6 = 1$, $X_3 = X_7 = 0$.

**7-26** (*b*) $X_7 = 1$, $X_3 = X_5 = X_6 = D$, all other $X = 0$.

**7-28** (*b*) Three transistors; *Q*0 and *Q*1 each have two emitters, *Q*2 has three emitters.

**7-33** $Y_1 = W_6 + W_7 + W_2\overline{W}_4\overline{W}_5 + W_3\overline{W}_4\overline{W}_5$.

**7-36** (*a*) 10; (*b*) 7.

**7-39** (*a*) $Y_5 = \overline{D}\,\overline{C}\,\overline{B}A + \overline{D}\,\overline{C}B\overline{A} + \overline{D}\,\overline{C}BA + \overline{D}CBA + D\overline{C}B\overline{A} + D\overline{C}BA + DCBA$.

**7-41** (*a*) 6; (*b*) 4.

**7-42** (*a*) 7; (*b*) 3; (*c*) 6, 4; (*d*) 1024.

## CHAPTER 8

**8-5** 0.3 V, 2.7 V, 533 mV.

**8-20** (*b*) One pulse, $N = 5$.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

You have either reached a page that is unavailable for viewing or reached your viewing limit for this book.

**TABLE 6-3 Summary of Basic Boolean Identities**

*Fundamental laws*

| OR | AND | NOT |
|---|---|---|
| $A + 0 = A$ | $A0 = 0$ | $A + \bar{A} = 1$ |
| $A + 1 = 1$ | $A1 = A$ | $A\bar{A} = 0$ |
| $A + A = A$ | $AA = A$ | $\bar{\bar{A}} = A$ |
| $A + \bar{A} = 1$ | $A\bar{A} = 0$ | |

*Associative laws*

$(A + B) + C = A + (B + C) \qquad (AB)C = A(BC)$

*Commutative laws*

$A + B = B + A \qquad AB = BA$

*Distributive law*

$A(B + C) = AB + AC$

*De Morgan's laws*

$\overline{AB\cdots} = \bar{A} + \bar{B} + \cdots$

$\overline{A + B + \cdots} = \bar{A}\bar{B}\cdots$

*Auxiliary identities*

$A + AB = A \qquad A + \bar{A}B = A + B$

$(A + B)(A + C) = A + BC$

**TABLE 8-2 FLIP-FLOP Truth Tables**

| *SR* | | | *J-K* | | | *D* | | *T* | | *Direct inputs* | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $S_n$ | $R_n$ | $Q_{n+1}$ | $J_n$ | $K_n$ | $Q_{n+1}$ | $D_n$ | $Q_{n+1}$ | $T_n$ | $Q_{n+1}$ | *Ck* | *Cr* | *Pr* | *Q* |
| 0 | 0 | $Q_n$ | 0 | 0 | $Q_n$ | 1 | 1 | 1 | $\bar{Q}_n$ | 0 | 1 | 0 | 1 |
| 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | $Q_n$ | 0 | 0 | 1 | 0 |
| 0 | 1 | 0 | 0 | 1 | 0 | | | | | 1 | 1 | 1 | * |
| 1 | 1 | ? | 1 | 1 | $Q_n$ | | | | | | | | |
| Fig. 8-9 | | | Fig. 8-11 | | | Fig. 8-14 | | Fig. 8-15 | | | | | |

*Refer to truth table *SR*, *J-K*, *D*, or *T* for $Q_{n+1}$ as a function of the inputs.

Tata McGraw-Hill
Publishing Company Limited
7 West Patel Nagar, New Delhi 110 008

*A Division of The* ***McGraw-Hill*** *Companies*

